Table of Contents

Preface 1
Conference Overview 3
Conference Program 5
Introduction 8

PEER REVIEWED PAPERS

A case study of tensions in student-faculty partnerships for departmental change work
  Fatima N. Abdurrahman, Diana Sachmpazidi, Robert P. Dalka, and Chandra Turpen 10

Analyzing the functions of multiple external representations of electric potential
  Jonathan W. Alfson, Paul J. Emigh, and Elizabeth Gire 16

Understanding students’ struggles with collaboration through their views of knowing
  Josephine R. Allen, William Henriquez, Thanh K. Lê, Andrew Boudreaux, and Carolina Alvarado 22

Air particles in a lattice: Considerations for sound wave simulations in physics education
  Jared Arnell and Hillary Swanson 28

Modeling confusion in collaborative learning
  Peter Bagdovitz and Jayson M. Nissen 34

“Science happens between people”: teachers’ perspectives in a physics RET program
  Roshni Bano and Minjung Ryu 40

Acquisition of qualitative video data: methods and reflections in PER
  Lauren A. Barth-Cohen, Tamara G. Young, Jason M. May, and Adrian L. Adams 46

Shifts in students’ responses to conceptual questions after a new physics conceptual worksheet: Preliminary findings
  Lauren C. Bauman, Lisa M. Goodhew, Anne T. Alesandrini, Al K. Snow, and Amy D. Robertson 52

Analysis of student essays in an introductory physics course using natural language processing
  Amir Bralin, Jason W. Morphew, Carina M. Rebello, and N. Sanjay Rebello 58

Investigating the Assessment Landscape of Physics Graduate Programs
  Bill Bridges, Caitlin Solis, Joshua Barron, James T. Laverty, Jacquelyn J. Chini, and Rachel Henderson 64

Investigating context dependence of introductory and advanced student responses to introductory thermodynamics conceptual problems
  Mary Jane Brundage, David E. Meltzer, and Chandralekha Singh 70

Considering the Departmental Action Leadership Institute as a Community of Transformation: What’s highlighted and what’s missed?
  Robert P. Dalka, Chandra Turpen, Diana Sachmpazidi, Fatima N. Abdurrahman, David A. Craig, and Joel C. Corbo 76

Using clusters of models of disabilities to describe support for mentees with disabilities
  Constance M. Doty, Dan P. Oleynik, Erin M. Scanlon, and Jacquelyn J. Chini 82
Student belonging in STEM courses that use group work
Paul J. Emigh, Sujata Krishna, Jiehong Liao, Katsuhiro Kita, Jennifer R. Casey, and Jayson M. Nissen 88

Learning assistants’ teaching strategies for promoting scientific inquiry among undergraduate students in a physics laboratory setting
Samuel W. Engblom, Maggie S. Mahmood, Michael J. Matos, and Jackie Vargas 95

Applying Voting Theory to Mastery Grading: A Study of Faculty Interpretation of Course-Level Categorical-Score Distributions
M. T. Freeman, Amogh Sirnoorkar, James T. Laverty, and Bethany R. Wilcox 101

Broadening Student Learning through Informal Physics Programs
Carlee Garrett, Tatiana L. Erukhimova, Jonathan D. Perry, and Jonan Phillip Donaldson 108

Representational differences in how students compare measurements
Gayle Geschwind, Michael Vignal, and Heather J. Lewandowski 114

Effect of Essential Skills Practice on Student Understanding of Probabilities in an Upper-Division Quantum Mechanics Course
John Goldak and Peter S. Shaffer 120

Relating visual attention and learning in an online instructional physics module
Razan Hamed, N. Sanjay Rebello, and Jeremy Munsell 126

A Look Physics Teacher Identity Around Equitable Instruction: The Tour Guide, Coach, and Gardener
Maria S. Horak, Clausell Mathis, Delwrick Nanthou, Abigail R. Daane, and Michelle N. Brown 132

Development, validation and online and in-person implementation of clicker question sequence on change of basis
Peter Hu, Yangqiuting Li, and Chandralekha Singh 138

Simulated Groupwork in an Asynchronous Course Learning about Radioactivity
Michael M. Hull, Maximilian Jeidler, and Eva Holzinger 145

Drawing on force ideas for kinematic reasoning in introductory physics
Tra Huynh, Anne T. Alesandrini, Lauren C. Bauman, Olin Sorby, and Amy D. Robertson 151

Toward helping students develop error detection skills
Safana Ismael and Mila Kryjevskaja 157

Students Stumble With Inconclusive Results: An Exploratory Analysis on How Students Interpret Unexpected Results
Joss Ives, Aaron M. Kraft, James Day, and Doug A. Bonn 163

Research on a faculty support program for working with learning assistants
Adrian Juanson, Gina M. Quan, Jennifer S. Avena, and Alexandra Gendreau Chakarov 169

Rasch Analysis of the Quantum Mechanics Concept Assessment
Jesse Kruse and Bethany R. Wilcox 175

Metacognitive knowledge and regulation of peer coaches
Jonathan Kustina and Dawn C. Meredith 181

Physics Teachers' Motivations to Learn Computational Thinking as a Re-novicing Experience
W. Brian Lane, Terrie M. Galanti, and XL Rozas 187
Exploring alternative perspectives through fictionalized student dialogues
Thanh K. Lê, Andrew Boudreaux, and Jayson M. Nissen

Adding self-regulated learning instruction to an introductory physics class
Danielle Maldonado and John Stewart

Upper-level students’ conceptual understanding of energy and momentum
Alexandru Maries, Mary Jane Brundage, and Chandralekha Singh

Augmented Reality to Scaffold 2D Representations of 3D Models in Magnetism
Michele W. McColgan, George E. Hassel, Natalie C. Stagnitti, Jason W. Morpew, and Rebecca S. Lindell

"Academia as a whole is structured entirely without any consideration for neurodivergency," and other things neurodivergent students want you to know
Liam G. E. McDermott and Nazeer A. Mosley

Evaluating learning of motion graphs with a LiDAR-based smartphone application
Colleen Megowan-Romanowicz, Daniel J. O’Brien, Rebecca E. Vieyra, Chrystian Vieyra Cortés, and Mina C. Johnson-Glenberg

How traditional physics coursework limits problem-solving opportunities
Barron J. Montgomery, Argenta M. Price, and Carl E. Wieman

Grappling with the dominant narrative of physics: Instructors rethink colonial roots together to reshape classrooms
Delwrick Nanthou, Ahmed A. Gumale, Clausell Mathis, Michelle N. Brown, and Abigail R. Daane

Supporting first- and second-order departmental change with the Effective Practices for Physics Programs (EP3) Guide
Christine O'Donnell and Stephanie Viola Chasteen

Three axes for expressing disability models and experiences: The cause, the effect, and the ability/disability dichotomy
Dan P. Oleynik, Constance M. Doty, Erin M. Scanlon, and Jacquelyn J. Chini

"The prettiest photos are the ones that have happy people in them": the use of photovoice in an upper-division physics capstone project course
Kristin A. Oliver, Victoria Borish, Bethany R. Wilcox, and Heather J. Lewandowski

Alignment between student epistemological views and experiences with course structures in introductory physics: A case study
Ellen Ouellette, Sarat Lewsirirat, Ryan Biju Sebastian, Morten Lundsgaard, Christina Krist, and Eric Kuo

"Which has more energy?" - An example of responsive teaching in university physics
Jon C. Owen, Lisa M. Goodhew, and Brynna Hansen

Students’ interpretations of disciplinary convention with the first law of thermodynamics
Alexander P. Parobek and Marcy H. Towns

LA Program as a Driving Force for Identity Development Through Access to Ideational Resources
Xandria R. Quichocho, Simone Hyater-Adams, and Eleanor W. Close

Tools for Understanding the Microscopic World of Quantum Mechanics: Analogies in Textbooks
Diana Ryder, Michael Verostek, and Benjamin M. Żwickl
Characterizing representational gestures in collaborative sense-making of vectors in introductory physics  
Stacy M. Scheuneman, Virginia J. Flood, and Benedikt W. Harrer

Physics Graduate Teaching Assistant Use of Error Framing in Recitations and Laboratories  
Daniel Sharkey, Constance M. Doty, Tong Wan, Erin K. H. Saitta, and Jacquelyn J. Chini

Analyzing students’ assumptions to varying degree of prompting during problem solving  
Amogh Sirnoorkar and James T. Laverty

Investigating how students engage with a digital planetarium  
Alexander Sivitilli and Saalih Allie

Unlearning indoctrination: Tensions between decolonizing curricula and characteristics of whiteness  
Mathilda J. Smith, Michelle N. Brown, Abigail R. Daane, and Clausell Mathis

A case in which canonically incorrect ideas do not hinder conceptual progress in introductory physics  
Al K. Snow, Paula R. L. Heron, Lauren C. Bauman, Amy D. Robertson, and Lisa M. Goodhew

Conceptual Challenges of Discretizing Wave Functions: A Case Study  
Christian D. Solorio and Elizabeth Gire

Using a volunteerism framework to understand the motivations of university students who facilitate informal physics programs  
Bryan Stanley and Kathleen A. Hinko

Defiance in the face of adversity: a qualitative study of women's attrition from and persistence in physics  
R. Smith Strain, Matthew Shepherd, and Eric Burkholder

Characterizing the 'design-science gap' in an engineering design-based laboratory unit in an introductory physics course for future engineers  
Ravishankar Chatta Subramaniam, Amir Bralin, Jason W. Morphem, Carina M. Rebello, and N. Sanjay Rebello

Investigating Academic Burnout in Undergraduate Physics Experiences  
Harshini Sunil and Bethany R. Wilcox

Testing for over- and under-dispersion in physics degree outcomes  
Astra Sword

Characterizing the complexities of experimental decision making in an introductory lab practical  
Hamideh Talafian and Katherine Ansell

Using Community Cultural Wealth to Understand Experiences in Physics Bridge Programs  
Jenna P. Tempkin, Geraldine L. Cochran, and Téa Boone

Inequities and misaligned expectations in PhD students' search for a research group  
Michael Verostek, Casey W. Miller, and Benjamin M. Zwickl

Investigating Perceptions of Relevance Towards Computation in an Introductory Physics for Life Science Course  
Jacob Watkins, Nikolay Ivanov, Kathleen A. Hinko, and Kirtimaan Mohan

ChatGPT reflects student misconceptions in physics  
Sav Wheeler and Rachel E. Scherr
Examining faculty choices while implementing the Next Gen PET curriculum through Revealed Causal Mapping
Julia Willison, Erin M. Scanlon, and Jacquelyn J. Chini

Analyzing the dimensionality of the Energy and Momentum Conceptual Survey using Item Response Theory
Xian Wu, Matthew W. Guthrie, Erin M. Scanlon, and Yaoguang Li

Students’ use of symmetry as a tool for sensemaking
Tamara G. Young, Lauren A. Barth-Cohen, and Sarah K. Braden

Analyzing Physics Majors’ Specialization Low Interest Using Social Cognitive Career Theory
Dina Zohrabi Alaee, Keegan Shea Tonry, and Benjamin M. Zwickl

Analyzing AI and student responses through the lens of sensemaking and mechanistic reasoning
Dean A. Zollman, Amogh Sirnoorkar, and James T. Laverty

List of Contributed Poster Presentations

Index
Preface

The theme of the 2023 PERC was: Working together to Strengthen the PER Community of Practice.

The 2023 Physics Education Research Conference (PERC) revisited the first PER community gathering that occurred nearly 30 years ago where a true PER community of advisors, professors, and students gathered. PERC 2023 conference was an invitation for the PER community to explore emergent themes in PER as a scholarly endeavor and as a research community. This theme was designed to encourage the continued formation of a community of practice as PER community evolves. In addition to the papers addressing this year’s theme, the remainder of the papers represent the diversity of current research within PER and help this volume fulfill its purpose of providing an annual snapshot of the field.

This year, the PERC took place in an in-person format. It is impossible to overstate the efforts of the organizing team, which included Rebecca Lindell, Liam McDermott, Jason Morphew, Mary Urquhart, and Lyle Barbato, who worked to create a meaningful conference experience and inclusive environment for all participants. The committee was supported in their efforts by leadership from both the American Association of Physics Teachers (AAPT) and the PER Leadership and Organization Council (PERLOC).

Plenary speakers included Ximena C. Cid, Kim Coble, Nekeisha Johnson, W. Brian Lane, Kris Lui, Jose P. Mestre, Mary Urquhart, Dean A. Zollman. The conference also hosted a variety of workshops, symposia, collaboration spaces, dine-n-discover events, in addition to the PERC poster sessions.

The 2022 PERC Proceedings process would not be possible without the ongoing support of Lyle Barbato, who works closely with the editors each year. The continued sponsorship of AAPT and open-access publishing through ComPADRE have allowed the PERC Proceedings to improve each year, including maintaining the double-confidential review process and continue the development of a robust and dedicated online platform for proceedings management.

However, it is the referees who volunteer their time and expertise each year who are critical to the PERC Proceedings success. The editors sincerely thank all of those who participated in the review process and strengthen our research community: Fatima Abdurrahman, Adrian Adams, Anne Alesandrin, Jonathan Alfson, Josephine Allen, Winter Allen, Saalih Allie, Carolina Alvarado, Katherine Crimmins, Jared Arnell, Jennifer Avena, Peter Bagdovitz, Roshni Bano, Lauren Barth-Cohen, Lauren Bauman, Scott Bonham, Tea Boone, Victoria Borish, Andrew Boudreaux, Sarah Braden, Amir Bralin, Bill Bridges, Mary Brundage, Eric Burkholder, Kevin Cantlin, Stephanie Chasteen, Ravishankar Chatta Subramaniam, Jacquelyn Chini, Eleanor Close,

By continuing to honor the PERC Proceedings tradition of welcoming new members in the field to publish their research alongside known scholars, it is our hope that the Proceedings can contribute in some small way to the goal of strengthening and improving the community of practices of PER for the years to come.

Qing X. Ryan
PERC 2023: Working Together to Strengthen the PER Community of Practice

Conference Overview
The theme of the 2023 PERC was "Working Together to Strengthen the PER Community of Practice." A community of practice, by definition, is a collection of individuals who "share a concern or a passion for something they do and learn how to do it better as they interact regularly." When the first PER community gathering occurred nearly 30 years ago, it represented a meeting of a true PER community of practice, made up of advisors, professors, and their students. Since then, the PER community has evolved and now consists of a much broader and diverse community.

The PERC '23 organizing team believes the PERC needs to evolve in order to become a true community of practice for the now broader and more diverse PER community. Our vision of this theme culminated in a conference that allowed members of the community to gather, discuss, share, learn, and work together to improve the field of Physics Education Research as a whole. We thank our diverse community that gathered and collaborated with mutual respect.

Organizers
Rebecca Lindell, Tiliadal STEM Education: Solutions for Higher Education
Liam McDermott, Rutgers University
Jason Morphew, Purdue University
Mary Urquhart, University of Texas at Dallas
Lyle Barbato, American Association of Physics Teachers (Volunteer)

Acknowledgments
The PERC 2023 organizers deeply appreciate everyone who worked to make this conference successful, including the:

Plenary Speakers:
Ximena C. Cid, California State University - Dominguez Hills
Kim Coble, San Francisco State University
Nekeisha Johnson, North Dakota State University
W. Brian Lane, Co-Director of Northeast Florida Center for STEM Education, Instructor of Physics
Kris Lui, OPTYC and the American Association of Physics Teachers
Jose P. Mestre, University of Illinois, Urbana-Champaign
Mary Urquhart, University of Texas at Dallas
Dean A. Zollman, Kansas State University

Session Organizers, Moderators, and Facilitators:
Scott E. Allen, Abigail Daane, Frank Dachille, Danny Doucette, Matthew Guthrie, Charles Henderson, Paula Heron, Dena Izadi, W. Brian Lane, Rebecca Lindell, Kristine Lui, Liam McDermott, Sam McKagan, Gina Quan, Erin Scanlon, Christian Solorio, J. Caleb Speirs, Tim Stelzer, Fatih Tasar, DJ Wagner, and Xian Wu

**Community Connection Hosts:**
Kim Coble, Dedra Demaree, Charles Henderson, Rebecca Lindell, Liam McDermott, Jason Morphew, Erin Scanlon, Mary Urquhart, Gary White, and Raymond Zich

**AAPT staff:**
Cerena Cantrell, Tiffany Hayes, and Jamar Jennings

**PERLOC Liaisons:**
Raymond Zich and Sujata Krishna

**PERC Proceedings Editors and Staff:**
Dyan Jones, Qing Ryan, and Andrew Pawl

**Financial Support:**
PERLOC and the AAPT

Finally, the PERC 2023 organizers would like to thank the attendees who all came to the conference and engaged with openness and grace.
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>2:00pm</td>
<td>Bridging plenaries and PERC Kickoff</td>
</tr>
<tr>
<td></td>
<td>The birth and early growth of the Physics Education Research Conference</td>
</tr>
<tr>
<td></td>
<td>Speaker: Dean Zollman - Presentation</td>
</tr>
<tr>
<td></td>
<td>The PER Community of Practice: The seen and unseen</td>
</tr>
<tr>
<td></td>
<td>Speaker: Ximena Cid - Presentation</td>
</tr>
<tr>
<td>4:00pm</td>
<td>Community Connections 1</td>
</tr>
<tr>
<td></td>
<td>Alternative Careers for PER Graduates: High School Teachers</td>
</tr>
<tr>
<td></td>
<td>Host: Dedra Demaree (she/her), Blue Ridge School</td>
</tr>
<tr>
<td></td>
<td>Connecting PER to K–12</td>
</tr>
<tr>
<td></td>
<td>Host: Mary Urquhart (she/her), University of Texas at Dallas</td>
</tr>
<tr>
<td></td>
<td>Disabled Physicists Meetup</td>
</tr>
<tr>
<td></td>
<td>Host: Rebecca Lindell (she/her), Tiliadal STEM Education: Solutions for Higher Education</td>
</tr>
<tr>
<td></td>
<td>Meet with your Journal Editor I: Physical Review PER</td>
</tr>
<tr>
<td></td>
<td>Host: Charles Henderson (he/him), Lead Editor, Physical Review Physics Education Research, Western Michigan University</td>
</tr>
<tr>
<td></td>
<td>New/Unconnected Researcher Connection, Collaboration, and Creation Space</td>
</tr>
<tr>
<td></td>
<td>Host: Liam McDermott (he/they), Rutgers University</td>
</tr>
<tr>
<td></td>
<td>Solo PER Connections</td>
</tr>
<tr>
<td></td>
<td>Host: Raymond Zich, Illinois State University</td>
</tr>
<tr>
<td>4:45pm</td>
<td>Poster Session 1 Setup</td>
</tr>
<tr>
<td>5:00pm</td>
<td><strong>Poster Session 1</strong> Abstracts and layout</td>
</tr>
<tr>
<td></td>
<td>Presenters with odd-numbered posters will present during the first 35 minutes. Following a 5 minute transition period, presenters with even-numbered posters will present during the last 35 minutes.</td>
</tr>
<tr>
<td>6:15pm</td>
<td>Dine and Discover</td>
</tr>
<tr>
<td>8:30pm</td>
<td>Plenary 2</td>
</tr>
<tr>
<td></td>
<td>Speakers: W. Brian Lane, Mary Urquhart, Kris Lui, and Nekeisha Johnson</td>
</tr>
<tr>
<td>PDT</td>
<td>Thursday, July 20, 2023</td>
</tr>
<tr>
<td>-------</td>
<td>------------------------</td>
</tr>
<tr>
<td>7:00am</td>
<td>Mentoring Meet Up</td>
</tr>
</tbody>
</table>
| 8:00am| Parallel Sessions Cluster 1  
**Researching and Interrogating Grading Practices** *(Poster Symposium)*  
*Moderator: J. C. Speirs, W. B. Lane; Presenters: A. G. Cano, W. B. Lane, S. Nguyen, C. Paul, B. Pollard, M. Swartz; Preparatory Material: Smith and Smith; Paul and Webb; Simmons and Heckler*  
**Team-work in Labs: Evidence from Research** *(Talk Symposium)*  
**Success Stories in Education Reform** *(Talk Symposium)*  
*Moderator: C. Henderson; Presenters: S. Allen, S. Chasteen, S. Pollock, T. Stelzer*  
**Research-based Assessments in the Digital Age: Validity, Reliability, Accessibility** *(Round Table)*  
*Facilitators: D. Wagner; Preparatory material: Madsen and McKagan and Van Dusen et al.*  
**Doing PER in Partnership with Two-Year Colleges** *(Critical Conversation)*  
*Presenters: A. Daane, K. Diff, K. Lui, C. Monsalves, S. Savrda, V. Sawtelle; Preparatory material: Kanim and Cid and viewing AACC's 2023 Fast Facts*  
**Using MAXQDA Software Effectively and Efficiently for Qualitative and Mixed Method Data Analysis** *(Workshop)*  
*Presenters: D. Izadi, C. Myers, B. Stanley; Facilitators request participants have a laptop with the MAXQDA trial installed*  
| 9:30am| Break |
| 9:45am| Plenary 3  
**Empowering scholars to change the culture of STEM pedagogy to be more inclusive and equitable**  
*Speaker: Kim Coble - Presentation*  
**Reforming Large STEM Introductory Courses Through Communities of Practice**  
*Speaker: Jose P. Mestre - Presentation*  
**PERC Proceedings Timeline**  
*Speaker: Lyle Barbato - Presentation*  
| 11:15am| Community Connections 2  
**Alternative Careers for PER Graduates: Primary Teaching Positions at Large R1 Universities**  
*Hosts: Erin Scanlon (she/her), University of Connecticut – Avery Point; Xian Wu, University of Connecticut; Daryl McPadden, Michigan State University*  
**Intro Astronomy Community**  
*Host: Kim Coble, San Francisco State University*
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>12:00pm</td>
<td>Lunch on your own</td>
</tr>
<tr>
<td>1:45pm</td>
<td>Poster Session 2 Setup</td>
</tr>
<tr>
<td>2:00pm</td>
<td><strong>Poster Session 2</strong></td>
</tr>
<tr>
<td></td>
<td>Presenters with odd-numbered posters will present during the first 35 minutes. Following a 5 minute transition period, presenters with even-numbered posters will present during the last 35 minutes.</td>
</tr>
<tr>
<td>3:15pm</td>
<td>Break</td>
</tr>
<tr>
<td>3:30pm</td>
<td><strong>Parallel Sessions Cluster 2</strong></td>
</tr>
<tr>
<td></td>
<td><em>Referee Guidelines to maintain clarity and generativity of quantitative research studies in Physical Review Physics Education Research</em> (Critical Conversation)</td>
</tr>
<tr>
<td></td>
<td><em>Presenters: E. Brewe, A. Heckler, R. Henderson, E. Kuo, T. Stelzer</em></td>
</tr>
<tr>
<td></td>
<td><em>Computation in the Physics Classroom and Physics Curricula</em> (Poster Symposium)</td>
</tr>
<tr>
<td></td>
<td><em>Moderator: C. Solorio; Presenters: T. M. Galanti, W. B. Lane, T. O. B. Odden, X. Rozas, H. C. Sabo</em></td>
</tr>
<tr>
<td></td>
<td><em>The “International Handbook of Physics Education Research”: Insights, Themes, and Future Directions</em> (Round Table)</td>
</tr>
<tr>
<td></td>
<td><em>Moderator: F. Tasar</em></td>
</tr>
<tr>
<td></td>
<td><em>Comparing different models for instructional materials based on a resources framework</em> (Critical Conversation)</td>
</tr>
<tr>
<td></td>
<td><em>Moderator: S. McKagan; Presenters: A. Elby, E. Etkina, L. Goodhew, A. Robertson; Preparatory material: Etkina and Van Heuvelen; Boudreaux and Elby; Robertson et al.</em> and websites <em>Isle Physics</em>; <em>Open Source Tutorials</em>; <em>ACORN Physics Tutorials</em></td>
</tr>
<tr>
<td></td>
<td><em>Exploring Culture in Physics</em> (Round Table)</td>
</tr>
<tr>
<td></td>
<td><em>Discussant: A. Daane; Presenters: B. Archibeque, A. Chase, F. Dachille, C. Mathis</em></td>
</tr>
<tr>
<td>5:00pm</td>
<td><strong>Using AI to Streamline your Physics Education Research Practice</strong> (Workshop)</td>
</tr>
<tr>
<td></td>
<td><em>Facilitators: R. Lindell, L. McDermott</em></td>
</tr>
<tr>
<td></td>
<td>Facilitator request: Participants have a laptop with free <em>Chat GPT 3.5</em>, <em>Bard</em>, and <em>Bing AI</em> personal accounts.</td>
</tr>
<tr>
<td></td>
<td><strong>Fostering Effective and Inclusive Group Work</strong> (Workshop)</td>
</tr>
<tr>
<td></td>
<td><em>Facilitators: M. Guthrie, E. Scanlon, X. Wu</em></td>
</tr>
<tr>
<td></td>
<td><strong>End of PERC</strong></td>
</tr>
</tbody>
</table>
Introduction

The 2023 Physics Education Research Conference Proceedings consists entirely of peer-reviewed papers.

The peer-reviewed papers are written products of any conference presentation made during the parallel and poster sessions. Each paper undergoes a rigorous peer review process in order to be published in the Proceedings. This year there were 94 submitted manuscripts, of which 68 were accepted for final publication.

The readership of the Physics Education Research Conference Proceedings includes faculty, post-doctoral associates, and graduate and undergraduate students in physics education; scholars in other discipline-based science education or closely related fields, such as cognitive science; practitioners in physics or other sciences, such as teaching faculty at undergraduate and graduate levels, and high school physics teachers.
PEER REVIEWED PAPERS
A case study of tensions in student-faculty partnerships for departmental change work

Fatima N. Abdurrahman, Diana Sachmpazidi, Robert P. Dalka, and Chandra Turpen
Department of Physics, University of Maryland, 4150 Campus Dr, College Park, MD 20742, United States

Students as Partners (SaP) is a pedagogical approach that considers students co-creators of an educational environment along with faculty, rather than passive participants. While an increasing body of literature evidences a multitude of positive outcomes from the SaP approach, there remains limited research on the challenges that arise in such collaborations. Quan et al. (2021) outlined such challenges in a paper showing that different members of a Departmental Action Team (DAT), in which students, staff, and faculty collaborate on a change effort, had different perspectives of their partnership. In this work, we confirm and expand upon those findings in a case study of another DAT. Our case study DAT comes from the first cohort of the Departmental Action Leadership Institute (DALI), a workshop series that supports faculty members in physics departments facing major challenges or opportunities. We find that all points of disconnect from Quan et al. are present in our case study. Additionally, we identify three specific areas of differing perspectives between faculty and students: motivation, commitment duration, and information transparency. We present evidence of these tensions with interviews from faculty, student, and alumni DAT members. Finally, we discuss how these tensions may be navigated by faculty seeking to partner with students in departmental change work.
I. BACKGROUND

The Effective Practices for Physics Programs (EP3) Initiative aims to support undergraduate physics departments in change efforts via the EP3 Guide [1] and the Departmental Action Leadership Institute (DALI) [2]. First launched in 2021, the DALI hosts teams of physics faculty in a year-long series of workshops on sustainable institutional change and the facilitation practices needed to lead teams in such efforts. Each participating program sends a team of two "change leader" faculty members to the DALI, who concurrently establish and facilitate a Departmental Action Team (DAT) at their home institution [3, 4]. These DATs are composed of faculty, students, and/or staff, in addition to the two DALI-participating change leaders. These department members collaborate to address a broad departmental issue (e.g., retention, curriculum) for at least a year.

A central principle of the DAT model is that students are part of the educational process. This value of this principle is illustrated by the growing body of research on "students as partners" (SaP) [5]. SaP describes an approach for faculty to collaborate with students in the creation of educational environments, rather than see them as passive participants. Importantly, the collaboration features a "reciprocal process through which all participants have the opportunity to contribute equally, although not necessarily in the same ways, to curricular or pedagogical conceptualization, decision-making, implementation, investigation, or analysis" [6]. The literature on SaP reports a wide range of positive outcomes for both students and faculty, including building trust, increased engagement, and learning opportunities [7].

While there is increasingly strong evidence of the value of SaP in higher education, a systematic review of SaP literature identifies several aspects of the approach that require further study [5]. One such aspect was that of challenges in faculty/student collaborations, a greater understanding of which would "help to complete the narrative around these complex relationships." Indeed, while studies on the benefits of SaP are required to make a compelling case for the inclusion of students in departmental change, knowledge of the difficulties that arise in such partnerships can be invaluable to navigating them effectively. This utility is exemplified by Quan et al. [8], which unpacks challenges faced by students and faculty as they partner on a DAT. In doing so, they find students and faculty can hold different views of their partnership, and subsequently, have different views on the work as well as power dynamics. Our study complements Quan et al.’s exploration of challenges in DAT collaborations, identifying additional points of tension in student-faculty partnerships on DATs.

One thread of this project has focused on change leaders’ experiences with student partnership in their DATs, and has produced findings on the positive outcomes of these partnerships [9]. This paper continues that thread by examining tensions that arose in their partnerships, drawing from a second case study department in the same 2021 cohort of DALI change leaders. We note that while the tensions presented later in this work were present across all three case study departments in the cohort (including Hemlock University, the case study presented in [9]), for the sake of brevity, we will focus on only one exemplary case study department.

Maple College (pseudonym) is a state college in the Northeastern United States serving predominantly local, low-income communities. The undergraduate physics program, which exists within a larger physical sciences department, employs 5 faculty members, and graduates 2-4 majors per year. The two faculty change leaders, Morgan and Misha, are both tenured professors who have been at Maple for 16 and 7 years, respectively. They applied to the DALI with the goal of increasing the recruitment of physics majors, citing a greatly reduced budget and low math preparation as critical challenges. In addition to the two change leaders, their DAT was established with three more physics faculty members (including the department chair), a faculty member from Maple College’s department of education, two students and a recent alumnus of the program. This work draws on data from interviews with Morgan and Misha as well as a subset of the DAT members. Each change leader was interviewed three times over the course of their year in the DALI. The two students, the alumnus, and the education faculty member were each interviewed once following the conclusion of the DALI, approximately a year after the DAT was established. The only non-change leader members represented in this paper are Megan (a 4th year who graduated a week after her interview) and Mike (who graduated shortly after the DAT was established, but stayed on for a year as an alumnus). The dates and participants of our interviews are summarized in Table 1.

<table>
<thead>
<tr>
<th>Date</th>
<th>Data Collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>March 2021</td>
<td>First change leader interviews</td>
</tr>
<tr>
<td>June 2021</td>
<td>Second change leader interviews</td>
</tr>
<tr>
<td>February 2022</td>
<td>Third change leader interviews</td>
</tr>
<tr>
<td>April/May 2022</td>
<td>DAT member interviews</td>
</tr>
</tbody>
</table>

Interviews were approximately 1 hour long each and conducted over Zoom with semi-structured protocols. Interview questions centered experiences on the DAT (e.g. "How are difference of opinion resolved on the DAT?", or "What roles have students played in the decision making process?") and expectations of the change effort ("Is there anything you think could impede your team’s success?"). The recordings of these interviews were transcribed and reviewed for segments dis-
cussing the partnership between students and faculty. From these segments, we identified points of tension between student and faculty experiences, either via contradicting perspectives (e.g., faculty report X while students report Y) or one party’s espoused confusion at the other party’s behavior (e.g., faculty member doesn’t understand why students do Z). In considering these discrepancies in perspectives, we draw on Quan et al.’s use of crystallization as an analysis perspective, assuming that differing perspectives between different members are simultaneously true, each representing different facets of the partnership [10]. Below, we outline three of these tensions and how they came up in Maple’s DAT.

III. FINDINGS

We find that all three points introduced by Quan et al. ("members of a partnership can hold different perspectives about the partnership", "members attend to different aspects of their work when thinking about partnership", and "power dynamics are more and less visible depending on roles") are present in our case study of Maple College’s DAT. To continue this exploration of ways in which students and faculty’s perspectives differ on change work partnerships, we share three additional tensions identified in our case study.

A. Members of a partnership are motivated by different factors

One discrepancy between student and faculty expectations is motivation. Early on, Misha (change leader) expressed an assumption that she would be highly motivated in the DAT’s work compared to students, who are "not gonna think all day about that." presumably because of their other commitments. However, by the third interview, Misha revisited this assumption, expressing surprise to have found the opposite to be true:

*Oddly enough, the students are very enthusiastic. They seem to be more interested, which is kind of weird because we thought that the faculty would be more interested because this is primarily a faculty job to do...We can’t tell the students to, do a lot of work in addition to everything they have, but faculty, we’re supposed to do this work. So I don’t know, but the faculty seem to miss more meetings and stuff.* —Misha, Interview 3

Here, Misha reveals her expectation that the central motivation for faculty’s involvement in the change effort: it is "a faculty job" which they’re "supposed to do". She contrasts this with students, whose participation is voluntary, and therefore, surprising when they exhibit a high level of motivation. She mentions here that students are more “interested” than faculty, and later expounds on this, reporting students as having better attendance and being more responsive than faculty. She claims this is the case even for alumni members, who, according to Misha "have no skin in the game," again, emphasizing their lack of obligation towards the change effort.

While faculty members are largely motivated by a sense of duty to their jobs and subsequently expected low motivation from students, student members communicated a wide range of motivations for their involvement, a central one being their desire to seize the rare opportunity to improve the educational experiences of future students:

*[Participating on the DAT] was difficult, because I am a full-time student and I have a full-time job outside of campus...but at the same time, I felt good that I was doing this because it felt like I was, in a way, putting my name out there with all these other faculty, alumni, or teachers, but I was hoping the future generation of college like, improve their program, making their journey a little bit easier than mine was and maybe have them go all the way through without quitting. So it was both a lot of pressure but also felt very good.* —Megan

Here, Megan (a student) conveys the faculty expectation (that students are very busy and have other concerns), but emphasizes that she nonetheless is highly motivated to contribute to the work of the DAT. In addition to being known by faculty, her contribution is driven largely by her own experience in the program and the possibility of helping students who will come after her. While this is generally students’ central motivation on DATs, Megan shares a number of other motivations, including the acquisition of institutional knowledge that is helpful to her and her peers (sometimes called "navigational capital" [11]), and making a good impression and connections with faculty members.

An area where this disconnect between students’ self-reported high motivation and faculty’s expectations of students’ low motivation arose was in faculty’s negotiations of student role on the DAT, which, as Misha demonstrates in the segment below, is informed largely by faulty expectations rather than students’ reality:

*We have to be careful of [overburdening students] because they are not getting anything in return for this effort that they’re putting in. They are graduating in a year or two, or they already graduated, so they’re not really getting anything out of this. So, it really doesn’t seem okay to ask more than one or two hours a week of them, because it’s really a commitment, and even showing up for a meeting every two weeks every hour, like an hour every two weeks is, it’s a lot of work for that. So we want to be mindful of not overburdening them.* —Misha, Interview 3

Here, Misha’s approach to involving students (what she feels she can ask of them as a time commitment) is informed by an expectation of students’ motivation, which flattens the
diverse benefits students themselves have reported receiving from their DAT work.

B. Members are involved on different timescales

Another point of tension between faculty and students arose from their different durations in the department. On one side, faculty members spend many years in their positions, and as a result, exhibit a measured, long term approach to thinking about the change effort. Morgan (change leader) spoke to a perceived shortcoming of student partnership given the faculty’s long-term commitment:

_One thing we’re kind of worried about on curriculum decisions, ...we’re not really sure that the students ... should be totally dictating what we do in the physics program because [the faculty] have to live with it for a lot longer than they do._ —Morgan, Interview 2

Morgan having to "live with" the outcomes of the DAT illustrates a fundamental difference between students and faculty in a partnership: while students can participate in the short term within their college time, faculty will potentially be stuck with the obligations and outcomes of the change work for much longer. As a result, faculty believe they will ultimately carry the bulk of the workload long term, with Misha noting this difference in involvement duration and the subsequent fact that in the end, "faculty will have to implement whatever it is that we decide".

Conversely, students demonstrate a higher sense of urgency, along with a frustration at faculty’s lack thereof:

_One thing that I would like [the faculty] to change is maybe a sense of urgency. I think that they’re planning everything for the future, like they just keep saying in the future, in the future, in the future. And I appreciate that they’re even trying, but I think a sense of urgency would be more appreciated._ —Megan

Megan illustrates here the attitude that comes with students’ inherently short-term participation on DATs: as they have a limited time to participate, they are disillusioned by faculty’s emphasis on a future they may not be around for.

While this disconnect between students’ and faculty’s sense of timeliness created concern in both parties for the pace of the other, both parties also revealed glimmers of sympathy for the other’s position that could help mitigate negative outcomes of this tension. On the student side, Megan conceded after the above segment that changes involving, for example, department budgets, take significant amounts of time and are therefore solely in the realm of faculty’s ability. This reframing of the faculty’s lack of "urgency" positions faculty as able to make changes students can’t because of their limited time in the department. Similarly, Morgan revisited her opinion of how much student input should inform long term changes in her final interview. Reflecting on the value of student input over the first year of the DAT, she says:

_It really made us start thinking about, oh maybe there are some real problems or real missed opportunities maybe would be a better way to put it in our curriculum, that would both make our current students feel more comfortable, and hopefully, then make a whole program more attractive to prospective students._ —Morgan, Interview 3

This establishment of a through-line between current students and future students reframes student input from a fleeting, transient perspective faculty will be stuck with implementing, to a perspective representative of a lineage of students that, collectively, exist in the department for just as long as any one professor, if not longer.

C. Members share and withhold information differently

A final point of tension between student and faculty DAT members is the differences in their expected transparency with regard to information. While there is some common experience between faculty and students in a department that leads to a body of shared information, each group has unique information as well, specific to their position and experiences.

For example, student members bring unique knowledge of students’ needs, obligations, and challenges, which change leaders on the DAT cite as one of the most valuable aspects of the DAT model. Morgan speaks to how this unique student knowledge was essential in many aspects of the data collection process:

_[Students] were really helpful with...the whole focus group planning. [They] focus in on the questions that they felt would be most important... They did have pretty strong opinions about what should be asked in that focus group and also the best ways to make the other students feel comfortable and open up and what they thought people would feel comfortable with and what they wouldn’t...They also just bring a different perspective in terms of...things that they’re curious about and think are important that we just were like, oh yeah maybe we should think about that._ —Morgan, Interview 2

This value of student knowledge comes up across change leader and other faculty members interviews, with students’ unique experiential knowledge playing key roles in determining what data to collect, how to collect it effectively, and how to analyze it, all necessarily based on their own experience as students. However, unlike student members whose willingness to share their unique knowledge with the DAT appears to be an assumed and central part of their role, faculty members are more guarded about what information they share. In
one segment, Morgan discusses this negotiation in terms of what is "appropriate" for students:

At some point we're wondering like, are there decisions that are perhaps not appropriate to involve the students in? But we haven't, we've saved deciding that...we will cross that bridge when we get to it. —Morgan, Interview 2

While Morgan notices this tension but leaves it open for future negotiation, Mike, who has been on the DAT both as a student and an alum and is therefore more able to see both sides, lays out the disconnect in its present state more frankly:

There's a lot of politics that goes on with the administration versus the department, and I don't know if it's necessarily appropriate to have the students see all of that. But not that I have graduated...it's easier to see all those things because I'm not, you know, involved in the school anymore. But as a student I think it was a little tough to see the bigger picture and see what we were able to do versus what we wanted to do...They did a good job with communicating it, but it was just, it was a struggle because you don't wanna give out too much information, because again, it's just students, you know you don't wanna you know trash talk administration. —Mike

From his vantage point having recently occupied different positions on the DAT, Mike is able to clearly articulate both the faculty concern of oversharing with students, and the student frustration of not being able to see "the bigger picture" due to a lack of information.

IV. DISCUSSION

Through this case study, we illustrate three points upon which students and faculty collaborating on a change effort have different perspectives. We note that, while we have shared data from a single case study department (Maple College), these three tensions arose in all three of the case study departments’ DATs from the DALI’s first cohort. Rather than see the ubiquity of these challenges as a shortcoming of the SaP approach, we believe that understanding the tensions inherent to a student-faculty partnership can help faculty members anticipate and/or navigate them when attempting to partner with students. As such, we wish to highlight how the differing perspectives that result in these tensions can be reconciled through early and direct communication.

For example, Misha initially expects that students, lacking the obligation for departmental work that faculty have, will have low motivation to participate on the DAT. However, she finds that students are more engaged than faculty; as Megan reports, students have a range of factors outside of job obligation that motivate their participation. This shows us how incorrect assumptions sometimes inform how faculty approach their partnership: Misha wants to, understandably, avoid overburdening students, but her assumption may be cutting short the extent to or ways in which students are willing (and even desire) to participate. However, this also shows us how collaboration with students gives faculty an opportunity to better understand them: eventually, the students’ high engagement challenged Misha’s expectations of them.

By understanding the reasons students participate in change efforts, their role in a collaboration could be determined by their actual motivations, rather than faculty expectations thereof. And while this understanding occurred organically in Maple’s DAT eventually, an early discussion of personal goals for involvement between team members could have avoid misperceptions between students and faculty. Similarly, the other tensions could also potentially be headed off by early, transparent discussions. For example, the point of students and faculty being in their departments for different durations (and therefore having different perceptions of a good or realistic timeline for change work) was the central complaint Megan had as a student on the DAT. However, an early conversation about anticipated timelines, constraints on time, and the difference between short and long-term change work could have alleviated this frustration.

We note that, in having these anticipatory conversations, faculty need not concede entirely to the student perspective; for example, in the case of the tension that arises from faculty withholding some information from students, there indeed may be types of information that are not appropriate for students. However, rather than let this close off the opportunity for conversation, it can be an opportunity for thoughtful examination of these boundaries. Consider: Is some information taken for granted as being inappropriate for students? Can some information be anonymized or otherwise altered such that it can be shared with students? What blindspots in students knowledge may hinder their ability to contribute to the partnership? These kinds of reflections can help reshape the boundary of shareable information in a way that increases team effectiveness while maintaining professionalism.

The EP3 Project research team hopes to continue our study of SaP in DALI-supported DATs. We are now collecting data from the third and fourth DALI cohorts, which will significantly increase our number of case studies and provide additional data for this analysis. We also hope to examine these case studies from different perspectives. Particularly, we plan to further investigate the experiences of students in these partnerships, identifying positive and/or negative outcomes they achieve through their participation.

ACKNOWLEDGMENTS

This work is supported by the NSF under Grant No. 1821372. We also thank the EP3 Research Team members and the study participants for making this work possible.


Analyzing the functions of multiple external representations of electric potential

Jonathan W. Alfson, Paul J. Emigh, and Elizabeth Gire

Department of Physics, Oregon State University, 301 Weniger Hall, Corvallis, Oregon 97331-6507

We present an excerpt from an in-class group activity where students generate equipotential curves for a quadrupole using: a whiteboard, a Mathematica notebook, and a 3D plastic graph. Applying Shaaron Ainsworth’s framework for the functions of multiple external representations, we analyze how the students used the three representations in concert. We found that the students used different processes for generating each representation. The highly complementary nature of the representations facilitated the group’s direct comparisons between representations, helping them to construct deeper understanding about the system and the representations. This case study also exemplifies a limitation of the Functions framework for multiple representations, namely that it does not consider the role of generating representations. We echo the calls to account for student generation in future analyses of the use of multiple representations, when relevant. 

Published by the American Association of Physics Teachers under a Creative Commons Attribution 4.0 license.
Further distribution must maintain the cover page and attribution to the article's authors.
I. INTRODUCTION & METHODS

Physicists use a variety of representations for solving problems and communicating concepts. These representations are sometimes used separately, but are often used in concert, as a set of multiple external representations (MERs). Much research has already studied the uses of various representations in physics; a subset of which investigates environments with MERs. (See, e.g., Refs [1–3].) The research presented here contributes to the existing body of literature about the use of MERs by applying Ainsworth’s Functions framework [4, 5] to a specific context: a single group of students working through an activity involving MERs (shown in Fig. 1). Although Ainsworth’s framework is widely cited, more papers demonstrating its use, particularly within physics and in classroom environments, are needed. This paper is intended to fill that gap.

The case study presented here is interpretive [6], because it will both illustrate the use of the representations through the Functions framework for MERs, and challenge some of the assumptions of the framework. For example, we will discuss how the students had to generate certain aspects of each representation, which is not addressed by the Functions framework. Ainsworth et al. [7] have recognized the need for more focus on multi-representational construction when thinking about how students learn with representations. In the work presented here, we found agreement with Ainsworth, Prain, Tytler, and others that future research should consider how students generate multiple representations. We believe this particular case is also interesting because the representations used have significant overlap in information and include a novel representation (the plastic graph). Our research question is: What facets of this group’s use of the MERs are captured by the Functions framework?

The instructional context of our research is an in-class activity that took place on the second day of an upper-division course on electrostatics in the Paradigms in Physics sequence, a set of reformed junior-level courses at Oregon State University [8]. For this activity, students were assigned to work in groups of three. Each group was prompted to draw on a tabletop whiteboard the equipotential curves due to four identical positive point charges at the corners of a square. After working together for some time to complete this part of the activity, the groups were directed to look at a pre-programmed Mathematica notebook displaying many different graphs of the electric potential for the system. Then, after some whole-class wrap-up discussion, the groups were given a second prompt: to sketch the equipotential curves for a new system consisting of a quadrupole square. During this time the groups continued to have access to the whiteboard and the Mathematica notebook and were invited to use a dry-eraseable 3D plastic graph, in which the height corresponds to the value of the electric potential in the plane of the quadrupole.

As part of a larger study of this activity, we examined video data of one three-student group (with pseudonyms Olive, Forest, and Sage, see Fig. 2). The larger study focuses on the relationship between each representation and the students’ science practices. In the process of analyzing that data, we identified moments where students were using the representations together. The focus of this paper is on some of those moments, which took place primarily in the second part of the activity. Author JA selected a portion of the transcript to specifically code for possible instances of each Function, and then discussed together with the other authors what those lines of code demonstrated. Author JA originally transcribed the video and has also periodically reviewed the video and transcript to ensure accurate summaries throughout.

II. THEORETICAL PERSPECTIVE

We use Ainsworth’s framework as the lens for studying the relationship between representations. This framework specifically addresses the use of multiple representations in terms of three different functions: complement, constrain, and construct.

Complement: MERs can complement each other by supporting different thinking processes or by containing different/partially redundant information about the same system. Processes contain strategies and tasks; a particular representation might be chosen to accomplish some task, or because it permits a particular problem-solving strategy.

Constrain: One (or more) representation(s) may also constrain the interpretation of some other representation(s), either by leveraging a user’s greater familiarity with a representation or due to the inherent properties of the representation(s). For example, some representations may be ambigu-
ous or unfamiliar to users and a more familiar representation can help with interpretation of the less familiar representation. Regarding inherent properties, we can consider the common practice of presenting a problem statement with both a diagram and a written description. The diagram constrains the written description to clarify a physical situation through the inherent properties of the diagram.

Construct: The third function of MERs is to help students construct deeper understanding via abstraction, extension, and relation. For this analysis, abstraction is how students use MERs to understand the essential elements of a concept itself, while extension and relation are the use of MERs to understand more about the representations themselves.

III. OVERVIEW AND DESCRIPTION OF THE ACTIVITY

In the first part of the activity, this group investigated the equipotential curves due to a collection of four positive charges on the corners of a square. The students drew curves on the whiteboard, considering the behavior close to single point charge as well as the spacing of the curves and behavior far away from the collection.

After some time working on this prompt, there was a whole-class wrap-up discussion where all the groups were directed to look at the pre-programmed Mathematica notebook and directly compare this with the whiteboard image. This notebook was available for the remainder of the activity.

The lead instructor then announced:

“So now I want you to make a quadrupole. So, the quadrupole is going to be two positive charges on opposite corners and two negative charges on opposite corners. Two positive charges and two negative charges. And I want you to draw the cross-sections of the equipotential surfaces for this plane.”

The group began drawing curves on the whiteboard and almost immediately modified the Mathematica notebook to represent the quadrupole. They then compared their whiteboard image and the Mathematica image. A few minutes later, the instructor announced that the plastic graphs were dry-erasable, and one of the students suggested that they should request a plastic graph and draw level curves on it. The group then drew level curves on the plastic graph and compared with their whiteboard image and the Mathematica image.

IV. RESULTS

We now describe the complement and construct functions of these three representations, and provide specific evidence for these two functions. Our data contains limited evidence of how any one representation constrains any other representation in this system, and we discuss this in Sec. VI.

A. Complementary Processes and Information

When this group studied the quadrupole, producing the three different representations of the equipotential curves each recruited different processes for reasoning about the shape. For the whiteboard image, the students first constructed some curves using their models for point charge electric potential. The students went through several brief reasoning steps for producing the equipotential curves; for example, they examined what happens very close to one of the point charges. They soon identified that, with the Mathematica notebook, they could change the signs of the charges to represent the new system:

Olive: “So then, we just need to change one of the negatives and one of the positives. So why don’t I just make this positive and make the other one negative [pause] right [pause] here.”

Mathematica allows precise scalar superposition at all points, which is not feasible with the whiteboard but is done for the students in Mathematica. After configuring the Mathematica notebook to represent the quadrupole, the students compared with and corrected their whiteboard image. They also used the Mathematica image to help them draw the rest of the curves on their whiteboard.

The students’ processes for generating curves on the whiteboard changed when the Mathematica notebook became available. Instead of using their physical reasoning to produce the curves on the whiteboard, the students simply referred to the Mathematica notebook. (Consider, for example, the first quote from Olive in Sec. IV B.) Forest recognized this at the very end of the activity, when he reflected:

Forest: “I also appreciate that we can successfully use technology to not have to think about stuff. I like that. [Olive and Sage nod]”

These students thus acknowledged the shift in process that occurred during this portion of the activity.

The plastic graph introduced another new process: drawing level curves and viewing the projection. This different approach influenced the students to request a plastic graph. When the instructor interrupted the class to announce that the plastic graphs are dry-erasable, Forest suggested to the group that they use one.

Forest: “Oh, shoot. Let’s do that.”
Olive: “What?”
Forest: “When we get the [plastic graph], let’s draw some rings on them.”
Olive: “Oh! Cool.”
Forest: “We can look at the projection.”

The students then asked for a plastic graph and went through the steps of drawing level curves on it and comparing it to the whiteboard.
FIG. 2. Olive, Forest, and Sage (left to right) drawing equipotential surfaces. Olive points out the difference between the whiteboard image and the Mathematica image.

There is a great deal of overlap in the information each representation contains. The similarity between the contours on Mathematica and the whiteboard image facilitated direct comparisons by the students. Initially, the plastic graph represents potential using height, but once the level curves were drawn, the representations are scaled in such a way that the students could directly compare by overlaying the plastic graph on the whiteboard. This overlapping information is connected to the other function we observed, constructing deeper understanding, which we discuss next.

B. Construct Deeper Understanding

Here we describe examples of the students using the representations for relation and abstraction. Extension is about using a familiar representation to learn something about an unfamiliar representation, but we did not see particular evidence for extension in our data set. No one representation was more useful for understanding the others—once the group had all three representations, they related them bi-directionally for each pair.

At several instances, the students compared the same feature of each representation to determine the similarities/differences. Once the group modified the Mathematica notebook to represent a quadrupole, Sage exclaimed,

Sage: “Yeah, I was right! [Points to computer screen.] On the asymptotes it’s zero because along those lines, there’s equal push/pull.”

Which Olive followed up with,

Olive: “Right. And then, yeah, so it is actually spaced farther out that way and closer this way [Pointing to the computer screen. See Fig. 2]. So it’s the opposite of what you [Forest] drew.”

We see in this interaction that the students immediately related what was on the computer screen with their whiteboard image and were particularly attentive to the lines of zero potential. They also discussed what the spacing of the equipotential curves looked like. This relation between the Mathematica notebook and the whiteboard continued through the rest of the activity.

When the group decided to get a plastic graph, they asked a learning assistant:

Forest: “Can we snag one? We’re trying to decide whether or not we think it’ll be fatter this way [toward the center of the collection] or fatter on the back end [on the outside of the collection].”

We see in this quote a plan to relate the plastic graph with the other representations, and to continue investigating the spacing of the curves.

The learning assistant gave a plastic graph to Forest and, once the group finished drawing some level curves together, Forest pointed out that there are straight line equipotentials on the plastic graph and commented on the spacing of the other curves,

Forest: “I mean you can definitely see though that there’s a line across this way [Draws straight line on plastic graph.], and you can see it, how they printed it, there’s a line across that way. [Draws perpendicular straight line on plastic graph.]

Olive: “Uh-huh.”

Forest: “So that’s clear. And those get fatter out that way. [Draws curve on plastic graph around negative pole. Curves from this interaction can be seen in Fig. 3]"

This interaction demonstrates the group’s efforts toward abstraction with regard to the lines of zero potential and the spacing of the oval-shaped curves. Moments after this, Forest sought to investigate outside the domain of the plastic graph,

Forest: “I’m just trying... I like our picture. I want to know what these do farther out... Is there a way?... Let’s do this. [Pulls up the laptop. Fig. 3]"

The group then worked together to expand the range of the Mathematica notebook to view a more zoomed-out image. By comparing Mathematica and their whiteboard image, the group constructed deeper understanding about both spacing and the lines of zero potential. Not only did the students relate representations, but they also used this relating to reveal the essential aspects of the equipotential curves.

V. DISCUSSION & CONCLUSIONS

Our research question is: What facets of this group’s use of the MERs are captured by the Functions framework? We
identified complementary processes and information. The processes that the students used were different depending on the representation they chose. Producing the curves on the whiteboard resulted in the students making some qualitative arguments about the shape of the curves. Using the Mathematica notebook allowed the students to modify an equation and apply superposition. The plastic graph let the students draw level curves and see a projection of those curves. These latter two representations helped the students make decisions about how to generate the whiteboard image. The information that each representation contains is very similar once produced (they are all images of level curves) and this resulted in a significant amount of direct comparison between representations. While Ainsworth considers different or partially redundant information to be complementary, we have seen that the students found value in the highly redundant information contained by these representations.

The students used these MERs to construct deeper understanding. The availability of MERs provided opportunities for connecting between representations, letting the students explore the system to grasp underlying patterns (abstraction). The Mathematica notebook and the plastic graph were each used to explore something that was observed on the other representation. The students chose to draw the equipotential curves on the plastic graph and placed it on top of the whiteboard, despite having already viewed these curves on the Mathematica notebook. Conversely, the students explored Mathematica when it was necessary to expand the domain, since the other representations have limited domains. Although the plastic graph was likely a new representation for these students (they had not seen it in class before and these are custom graphs), they did not use the whiteboard and Mathematica notebook to extend understanding for the plastic graph. Rather, they related all three representations. This indicates to us that the (novel) plastic graph did not significantly increase the amount of interpretation these students had to exercise.

Using the Functions framework, we have identified a distinction in the students’ processes due to complementing representations. Once the students had access to Mathematica for producing curves, they could use the software to generate the curves by choosing the appropriate signs of charges. With the plastic graph, the students could draw level curves and look at a projection. Our analysis also found that these students sought out the plastic graph precisely because of its potential for representing the level curves with high similarity to the whiteboard image. This implies that even representations with high degrees of similar information can result in constructing deeper understanding of a physical system, and students may even benefit from the highly similar nature. The implication for practitioners is that concerns about cognitive load from interpreting multiple representations may be minimal. Providing representations that support multiple processes engaged these students in forming deeper understanding of the physical system.

The representations are extremely similar—they are all 2-D curves—but each one is produced in a different medium using different reasoning and strategies. These considerations around generation are an interesting aspect of the complementary processes the students engage in, but it is difficult to describe with the Functions framework, because there is a tacit assumption that the representations have already been produced. Considering the affordances of representational construction [9, 10] appears to be a necessary part of understanding learning environments where students generate multiple representations.

VI. LIMITATIONS & FUTURE WORK

This study of in-class video has given us in situ information about the functions of these representations. Due to the in-class group setting, we did not gain much insight into each student’s familiarity with the representations, and this limited our evidence about the constraining function of the MERs. The high degree of similarity between representations also meant that constraint by inherent properties was not apparent in the students’ use of these representations. So we cannot make claims about the constraint function. A particularly valuable extension would be to conduct individual interviews to see more of how individual differences and inherent properties play a role in the constraining function of these representations.

VII. ACKNOWLEDGEMENTS

Aaron and Robyn Wangberg for their work on the Raising Physics to the Surface project, the OSUPER group, and the students who voluntarily participated in this research. Funded in part through the NSF DUE-1612480


Understanding students’ struggles with collaboration through their views of knowing

Josephine R. Allen (she/her/hers)
Department of Science Education, California State University, Chico, 400 W 1st St, Chico, CA 95929-0535

William Henriquez (he/him/his)
Department of Science Education, California State University, Chico, 400 W 1st St, Chico, CA 95929-0535

Thanh Lê (she/her/hers)
Department of Physics & Astronomy and Science, Math & Technology Education Program, Western Washington University, 516 High St, Bellingham, WA, 98225

Andrew Boudreaux (he/him/his)
Department of Physics & Astronomy and Science, Math & Technology Education Program, Western Washington University, 516 High St, Bellingham, WA, 98225

Carolina Alvarado (she/her/hers)
Department of Science Education, California State University, Chico, 400 W 1st St, Chico, CA 95929-0535

Student-centered learning environments are designed to support collaboration and exploration, directing learning into a collective experience. This case study explores a group of three students, who were highly vocal and determined to understand the activities. They showed attempts to engage in socio-metacognitive patterns, not always achieving it. For example, students explicitly communicated their need to work collaboratively, while another requested to work individually. This qualitative research project collected data from a physics undergraduate course for future k-8 teachers. We center the study on two students who were recorded during their classroom activities, participated in semi-structured follow-up interviews, and submitted reflections regarding their classroom experiences while collaborating in the same group. We analyze students' views of collaboration and reflections to understand their personal epistemology. We present how students' failed attempts at socio-metacognitive patterns can be understood through the differing perspectives of knowing they hold while collaborating as a group.
I. Introduction

Many student-centered, research-based physics curricula include an instructional strategy referred to as elicit-confront-resolve (ECR) [1, 2]. This strategy elicits common alternative thinking by posing questions known through research to be challenging for most students. Then, it confronts students with inconsistencies that arise from those answers, and finally, leads to students resolving those inconsistencies by building more normative understandings. Confusion often arises as students engage with elicit-confront-resolve activities. While confusion can lead to productive learning discussions, it can also lead to frustration [3].

Multiple studies have shown that, in ECR-based instruction, successful collaboration involves the active negotiation of shared understanding [5, 6]. Socio-metacognition refers to the ways a group monitors and regulates their interactions and collective learning processes. We have adopted the socio-metacognitive framework of Borge et al. [4] to gain insight into student engagement and learning during classroom activities that involve explicit confusion. This framework consists of six specific communication patterns surrounding collective information synthesis and knowledge negotiation. In this article, we focus on a single pattern, Developing Joint Understanding, in which a collaborative group ensures ideas are understood as intended by speakers by rewording, rephrasing, or asking for clarification.

In addition to the Borge framework for socio-metacognition, we have also drawn on theoretical ideas about students’ epistemological beliefs. We have found that examining individuals’ ways of knowing can support our understanding of the group interactions that occur during collaborative learning. Baxter-Magolda [7] has developed a framework for ways of knowing based on an extensive set of interviews with undergraduate students. In this framework, ways of knowing and patterns within them are socially constructed, and student activation of these patterns is fluid and context dependent. The framework includes four profiles, two of which were relevant for our study.

In the Absolute Knowing profile, the role of the learner is to obtain knowledge from the instructor; the role of peers is to share materials and explain what they have learned to each other; the role of the instructor is to communicate knowledge appropriately and ensure that students understand knowledge; and the role of evaluation is to show the instructor what was learned. In this profile overall, knowledge is regarded as certain or absolute.

In contrast, in the Independent Knowing profile, the role of the learner is to think for oneself, share views with others, and create one’s own perspective; the role of peers is to share views and serve as a source of knowledge; the role of the instructor is to promote independent thinking and the exchange of ideas; and the role of evaluation is to reward independent thinking. In the Independent Knowing profile, knowledge is regarded as uncertain, and each learner has their own beliefs.

In our work we are examining how individual students’ ways of knowing affect their negotiation of collective understanding during collaborative learning activities. This short paper addresses two specific research questions: How do students engage in the socio-metacognitive communication pattern of Developing Joint Understanding in this learning environment? What views of knowing do students hold while in this learning environment?

II. Methodology

Our research occurred in a physics course for preservice K-8 teachers at two universities on the West Coast, a Predominantly White Institution and a Hispanic Serving Institution. The course uses NextGen PET [8], a curriculum that includes ECR-based activities. A subset of the authors, who have taught the course many times over more than a decade, identified six lessons that commonly elicit confusion. During the 2021-2022 academic year, we collected video recordings of classroom interactions as well as students’ individual written reflections during these lessons. We also conducted individual out-of-class interviews with selected students.

In this article, we develop a qualitative case study of a group of three members: Adam, a White male student, and Angie and Belle, two White female students. This group was highly vocal and engaged in the lessons but struggled to collaborate. We identified moments of confusion from the classroom video and coded transcripts for instances of the Developing Joint Understanding communication pattern from the Borge framework [4]. Next, we analyzed transcripts of the interviews with individual students to better understand the individuals’ perspectives regarding their group’s interactions within this learning environment. Two of the three group members, Adam and Belle, participated in follow-up interviews. The interview questions included: How would you describe your relationship with science? What role or what impact does uncertainty and confusion have on your learning process in this course? In some lessons in this class, students may not immediately know whether or not their answer for a question is correct, and yet you are still expected to share your ideas. How has this format worked for you? We used Baxter-Magolda’s epistemological framework to identify students’ views about knowing in this learning environment [7]. In particular, the framework allows us to assess the extent to which different group members share consistent views about knowing.

Finally, we triangulate our analysis using Belle’s and Adam’s written reflections about their collaborative work. We reassess both their interactions and engagement in socio-metacognition considering their individual epistemological profiles.
III. Results

We present an overall view of the group engagement in socio-metacognition according to their classroom interactions; we then present data from their follow-up interview to explore their epistemological beliefs. We then present their individual reflections of their experiences in a laboratory session they shared.

A. Socio-metacognitive patterns

We observed Belle attempting to engage in collective information synthesis with her peers. Some of these attempts were successful, others were not. For example, after receiving clarification from her partner Angie, Belle notices Adam is confused. Belle offers to clarify the topic, without him explicitly prompting her for support. Noticing and responding to groupmates’ confusion is a common behavior implemented by Belle, which aligns with Developing Joint Understanding. She has a tendency to request support from her peers when unable to resolve her own confusion.

In contrast, we see Adam with a wider variety of approaches, which align with socio-metacognitive patterns sometimes while looking for individualized progress at other times. For example, the class structure often requires the students to use a whiteboard when presenting information to the class. Adam recognizes Belle’s need to create the whiteboard as a group, aligning with collective information synthesis, but also vocalizes his need to work independently first. Belle pushes, so Adam changes his approach to merge both individual and collaborative mediums at the same time, walking her through the process while doing it on his own. A different approach we see is him looking to resolve confusion in advance of the activity by reaching out to the instructors of the course rather than engaging in the activity with his peers.

B. Epistemological Assessment

We reviewed the full interview recording for both Adam and Belle, while identifying moments in which they alluded to the way they see learning, knowing, and collaboration in this learning environment. We present quotes from both of them organized by emergent themes from the analysis.

1. Views of science

Belle describes the positive impact of collaboration as a way to engage in science. It alludes to a process that values uncertainty and multiple attempts as part of it.

I have never liked science personally. I'm not the best at science and math, so I've just been standoffish from those subjects. I really like this class so far because learning is collaborative and learning is something that you should enjoy.

Adam mentions how he used to interpret scientific knowledge as the acquisition of facts and understanding.

I'll be honest, my relationship with science has actually fluctuated a little bit. When I [was] in high school, I started to have ideas... We have answers for everything and this is why I learned science. I'm learning about cells. I know exactly what A, B and C does and that's how that works.

Adam’s view has shifted to an understanding that scientific knowledge is not always absolute or certain; we as scientists are still working to understand scientific truth.

Now I feel I've actually surpassed that where I’m to this point where it's like, this is what I'm being taught, this is what the leading experts know, but there is still some convolutedness of this information up here. Even the top of the line scientists are still asking questions. [Interviewer interrupts saying we know a small amount]. Exactly. Having to understand that the world around you has an explanation, but also doesn’t.

Belle has a strong sense of science as a process while we see Adam moving from a set of facts to a process.

2. Engaging in Learning

Belle values collaboration as a key component of learning over individual engagement, showing frustration at being rushed through the process.

A lot of [group members] would just rush through the activities as if it was a race or that they didn’t even want to learn the material, that they just wanted to get this done. [...] It felt like a competition to get something done when that’s not what learning is. Learning is collaborative and learning is something that you should enjoy.

Adam describes the value of sitting in and working through your confusion during the learning process.

I think to have that initial, pondering questioning phase is really important because it opens your mind up to questioning, how does this work, without the answer, how does this work?

In contrast, 20 minutes later, he describes his own discomfort with this kind of engagement, preferring to know with certainty.

I get into this area where I'm like, am I right or am I wrong? The way I want to learn, the way I choose to push to learn. I want to learn things; I want to know things. Even though the culture of the class is like building blocks.

Belle values engaging with the process of learning, Adam also recognizes its value. In contrast, Adam also mentions his need to access knowledge with certainty at times.
3. The Role of Confusion or Uncertainty

Belle places a high value on working through confusion as part of the learning process to the point of describing it as a fun experience.

I think [uncertainty or confusion] makes me want to learn even more, because then that confusion will eventually lead into learning, which is really fun to-- get, I guess. [...] The point is to question and to have that confusion, and then it's what you do with that confusion and how you ask questions is really valuable.

Adam has discomfort in not knowing, preferring to skip the uncertainty in order to reach a resolution.

I'm a very hyper-fixated person. [...] Because in 90% of this course, we get presented with physical problems [...] We don't necessarily know the answer to them right away, and so my hyper fixed mind is like, "Okay. I want to get to the bottom of this, as just as possible. Because I want to know this." It bothers me when I don't understand how things work.

We notice how Belle values confusion as part of the learning process while Adam is "bothered" by it.

4. The Role of Peers

Belle relies on subtle cues for her group members to notice and respond appropriately to receive help.

I feel like I [ask for help in a really specific way] if I'm really not getting it, but if it's-- I usually just do cues because I-- This is weird, but I think that people can tell if I'm struggling in a weird way. I'm like, "Okay, so..." and then someone would continue my sentence or help me to explain it a little bit deeper.

In contrast, Adam is not focused on monitoring and interpreting his groupmates' social cues for help, but on uncovering the correct answer.

My entire focus, and you can see that, is on understanding. It's like okay, what is truth? What is truth? That's wrong in the sense of what we're supposed to be doing. We're supposed to be learning about how other people learn, I guess, which is on me, completely. Because it's like, I'm not focusing on my other classmates, I'm focusing on the material, not the metacognitive approaches.

This shows a difference in the role of peers for each of them. For Belle, peers are a key component in her learning, while Adam’s process does not depend on others.

5. Leveraging Knowledge in a Group

Belle, who has not taken a physics course before, believes that participants do not need to have a full conceptual understanding to be a valuable participant.

One time I raised my hand, and I got the answer wrong, like WRONG. Then, Jenn was like, “This is why I never raise my hand.” I was like, “Ohh.” That's like what we've just been taught our whole lives, is that your answer needs to be right, and I'm like, no, it doesn't. It can be wrong, and that's how you learn.

Adam shows how transitioning from an individual to a group pace was a difficult shift for him, from his previous physics classes. Also, there is an assumption that there is an expected level of preliminary knowledge gained from prior schooling which also regulates their interactions.

It was a big learning curve for me, [having] to slow down and really be okay with taking the pace of my people at the table. Definitely a hard thing to learn and definitely something I didn't get from the very beginning [...] I just assumed people were as educated as I was, which was definitely a mistake.

Belle considers initial ideas are useful to build an understanding, disregarding its accuracy. Adam is now aware of the different levels of physics knowledge people come into class with and sees it as a challenge to collaborate.

C. Individual written reflection on their experiences

We compare the individual written reflections from Belle and Adam. After collaborating in a lesson on energy transference models, students had some prompts to reflect on their experience. This is done individually after class.

The course encourages students to use whiteboards to share their ideas and results of certain sections of the work with the rest of the group. Belle shared “I found an understanding of [the topic] after doing most of the work and working through it on the whiteboard,” noticing this tool as a source of learning. Adam mentioned the use of whiteboards as a source to foment collaboration rather than understanding, “Whenever we completed a whiteboard we almost always included everyone as it was a very interactive part of the activity.”

The course is organized through students working in small groups during the lessons. We see how Belle’s interaction with a group gets compromised when it is seen as a competition rather than a collaboration:

With this past group that I have had it has been frustrating. Mainly because this class is a very collaborative class, my group has treated most of our work as a competition to be completed.

In contrast, when Adam is discussing the group, he focuses on the different paces the individuals have during the activity, which is not problematized:

When we were completing the activities, oftentimes some of us would fly ahead and complete more pages much more quickly than others in the group.

They both reflected on their group collaboration. Belle perceived it as frustrating given the lack of engagement in the process of understanding the phenomena:

While I make sure everyone is on the same page, most of my group seems as if they want to get the work done and checked off while I want to understand the work. As much as I want to get the work done too, I also want...
to make sure I comprehend it and in my group right now it just seems as if they don’t want to understand it, they want to get the points and move on.

In contrast, Adam perceived it as positive. He considered that each student had the space to contribute their ideas and were mindful of the other students’ learning process:

I think some of the most important aspects that helped facilitate strong group interactions was positive and respectful communication. This means allowing others to comfortably share their ideas, respecting the voice of others, and being mindful of your perspective on a situation as well as others.

Through these prompts, responded individually, we can observe how Belle’s focus was on the collaboration not supporting their understanding, while Adam’s focus was the ways of communicating rather than their collective understanding.

IV. Discussion

We observe that Belle’s engagement with this group is very consistent with the information shared during the interview and reflection. She considers learning to be accomplished through the process of collaboration between peers, who are seen as a resource in the learning process. There is no need for certainty in the learning process since she considers the exchange of opinions, given that each might have their own perspective, valuable. Therefore, we can see how she could be considered an Independent Knower in this context.

In contrast, we see a variety of ways in which Adam engages with the team, which is also reflected in his contrasting statements during the interview. On one hand, he states that knowledge is uncertain, even to the greatest scientists; he values allowing students to work through their confusion on their own and synthesize their own understanding. This aligns with him prioritizing Belle’s need for collaboration during the lesson even when he had requested time to work on it individually first. He recognizes that working through uncertainty is valuable for the learning process. These statements show Adam aligning with the views of an Independent Knower.

However, we also see Adam holding a different view of knowing. During the interview, he refers to his past as believing science was more align with a set of facts, to which he had changed. Nevertheless, he mentions being uncomfortable when knowledge is unknown, or understanding has not been attained. He states that it bothers him not understanding how things work, he wants to get to the answer or an explanation as quickly as possible. These statements align with him often requesting help from the instructors of the class, rather than seeing his peers as a resource. He usually engages in collaboration after attaining understanding as an individual. He recognizes a different pace in his learning progression between him and the rest of the tablemates, without perceiving it as a problem. Therefore, Adam also aligns with an Absolute Knower in this same context simultaneously.

The learning environment in which this class was set up requires students to collaborate through the experimental exploration of conceptual understanding. It pushes students to record and explore their initial ideas, confront them into a set of data collection, so they can come up with a supported view of the phenomena. Therefore, it fosters students engaging in socio-metacognitive patterns, especially Collective information synthesis. It provides the means for students to look for collective, rather than individual, understanding. Looking at this environment, we can see how it supports the success of Independent Knowers, as it views the nature of knowledge as uncertain and positions peers as a resource.

We believe that the differentiation between their epistemologies contributes to the group’s instances of unproductive collaboration. While Adam navigates this learning environment as an Absolute Knower in some instances and as an Independent Knower in others, it can pose challenges to a student who is consistently engaging as an Independent Knower. A collaboration between students that values the answer (Absolute Knower) more than the process (Independent Knower) can create a tension of goals during collaboration. When these differing perspectives are expected to develop a collective understanding through collaboration, differing needs will need to be met. During the classroom experience, both Adam and Belle are explicit about what they need from the other to succeed - Belle, collaboration, and Adam, in addition to moments of collaboration, individual thinking time. Because these epistemologies fundamentally differ from one another, there is an expectation that it will be challenging for the two students to adequately support the others’ learning needs while also supporting their own.

IV. Conclusions

In a learning environment that aligns with socio-metacognition, exploring students’ epistemologies can help us analyze students’ interactions. Increasing student awareness of their own epistemology could benefit their ability to engage in productive group collaboration. Students’ awareness and reflection on their thinking about learning could impact their future teaching practice. Particularly, noticing how the current science standards include practices around collaboration and communication, we consider it of high relevance to foster this reflection process in pre-service teacher populations.

Acknowledgments

This project was funded through the National Science Foundation Grant DUE-2021547 and DUE-2021307.
Air particles in a lattice: Considerations for sound wave simulations in physics education

Jared Arnell, Hillary Swanson

*Instructional Technology and Learning Sciences, Utah State University, 1400 Old Main Hill, Logan, Utah, 84322*

Many existing sound simulations show air particles as semi-stationary, as if suspended in a lattice. While this portrayal is based mainly on practical considerations, it can inadvertently reinforce non-normative conceptions. We share our observations of middle school science students interacting with lattice-style sound simulations, attending to the effect such representations had on their conceptualizations of sound. We discuss the implications of these results for the implementation of digital simulations in lessons on sound waves and provide suggestions for instructors to frame these tools and guide students’ attention toward relevant aspects of the visualization.
I. INTRODUCTION

The advent of the Digital Age granted physics educators access to a useful and powerful new tool: the digital simulation [1–3]. Physics phenomena are often complex and even unobservable, but simulations can strip away irrelevant or distracting information and make hidden events visible [4]. These affordances are especially useful for novices who have not yet developed the discernment to know where to look when observing phenomena, or constructed reliable mental models for visualizing abstract mechanisms. Moreover, simulations can be designed to connect with students’ prior knowledge and lived experience in order to address common misunderstandings [3]. When simulations are engaging, interesting, and designed with learning objectives in mind, they can add a rich new dimension to traditional instruction [1].

For all their utility, digital simulations are not a silver bullet and present several key challenges. Novices tend to interpret visual representations as literal analogs of the physical system, like seeing crossed lines on a graph as a physical intersection [5]. Naive learners can also trust simulations to a fault, assuming the experts who designed them could not be mistaken. This can even lead to students interpreting erroneous simulation data (caused by bugs or malfunctioning programs) as legitimate and meaningful [1]. On top of these risks, instruction that seeks to simplify or reduce the complexity of natural phenomena (as most simulations do) can inadvertently misrepresent the physics at play [6]. Given these hazards, the design and implementation of digital simulations in physics education must proceed with care.

Understanding sound requires connecting a microscopic model of particle interactions to a macroscopic model of everyday experience. Previous work in the domain has identified a number of widespread non-normative conceptions [6–10]. Students routinely imagine air molecules to be more-or-less stationary, swaying back and forth as sound waves pass by [7]. Similarly, they may characterize sound as a domino-like series of collisions [11]. Students may also misinterpret sinusoidal wave diagrams that portray abstract measurements of sound (e.g. pressure) as literal pictures of the sound waves [6, 10]. The variety and nuance of these beliefs means that students who provide technically-correct answers may not have necessarily arrived at a normative understanding of the wider topic [10, 12].

Digital simulations of sound waves have been successful at improving students’ conceptualizations of the phenomenon [14]. An archetypal example can be found among the PhET Interactive Simulations [2, 13, 15] developed by the University of Colorado Boulder, shown in Figure 1. As with most simulations of this design, the air molecules oscillate in place without travelling around the room. Simulations such as these can provide an informative (and even captivating) visual for enhancing classroom instruction. However, our team observed a concerning trend when utilizing simulations for sound waves in a middle school science class; the semistationary, coupled nature of the particles encouraged and re-
III. SETTING

The data for this study come from one classroom of 31 eighth-grade students who were participating in an instructional module pertaining to sound waves. The students participated in physical demonstrations and activities using musical instruments and slinkies to learn the basic characteristics of frequency, wavelength, and amplitude. Next, the students were shown two digital simulations [18, 19] of a speaker instigating sound waves through particles of air. Finally, the students designed their own interactive simulation in NetTango [20] (a block-based interface to NetLogo [21], an agent-based computational modeling environment) to explore the way changes to pitch and volume would affect the behavior of the particles of the medium.

Our observations focus on three questions from the students' worksheet, which contained their predictions and reflections from the various activities. The first (Q1) was posed after the students observed the digital simulations and asked them to predict what would happen to the air particles if the speaker was louder. The remaining questions were given after the students had created their own NetTango simulation for sound waves. These questions asked the students to describe the simulation’s behavior when the volume (Q2) or the pitch (Q3) was altered.

To identify patterns in reasoning, the students' answers were categorized by two independent coders based on the common terms or phrases which appeared in their explanations. The coders then engaged in social moderation [22] until complete agreement was reached.

IV. OBSERVATIONS

The instructional module for the sound wave unit included links to two publicly available digital simulations of longitudinal sound waves travelling through air [18, 19]. Both were structurally analogous to Figure 1: the air particles were represented by a cloud of dots with a few particles highlighted in red. A speaker pushed the nearest air particles, and each particle oscillated around an equilibrium point at their original position while transferring its motion to the neighboring particles. This caused a wave with dense crests and sparse troughs to propagate through the medium. The highlighted red particles showed how individual particles remained near their initial positions while the wave crest moved continuously across the room. The simulations were looped and could not be altered. The students were asked a series of questions relating to these simulations, including prompts about what each piece of the simulation represented, how the highlighted dots moved, and how energy was translated through the system. As a final prompt, Q1 gave the students a hypothetical consideration: "what do you think would happen to the air particles (dots) if the speaker was louder?"

Afterwards, the students used the block code in the NetTango modeling microworld to construct their own model of a speaker, medium, and listener as seen in Figure 2. The students added each agent to the model and specified their parameters. When the agents were deployed correctly, the simulation represented air particles as green dots which oscillated horizontally when perturbed. The students were challenged to alter the volume and pitch of the speaker and document the results. Q2 and Q3 asked for a description of the behavior of the particles when the volume or pitch was increased, respectively.

The most common responses, their frequencies, and a prototypical example of each can be found in Table I. It should be noted that not all students answered each question, and responses could belong to multiple categories. The faster responses explicitly referenced the increased speed of the particles. The movement responses mentioned a variety of changes in the motion of the particles, including that they would "go farther," "vibrate more," or move "more freely." The energy category included responses that explicitly mentioned "energy," as well as those which expressed similar notions like moving "harder" or "with more force." The frequency and amplitude categories counted responses which commented on the changing dimensions of the waves, such as being "closer together" or "bigger," respectively. The density responses described how the particles would be "more condensed" or "closer together." Finally, the more air responses stated that the number of air particles would increase.

Across all three questions, the faster category dominated the responses, showing up more than twice as often overall as the next most common category, movement. The movement, energy, frequency, and amplitude categories were seen in roughly similar amounts. The prevalence of frequency responses was much higher in Q3 than Q1 or Q2 due to the nature of the prompts; likewise, amplitude was more common in Q1 and Q2 than in Q3. Lastly, the density and more air responses were significantly less common than any other category.
TABLE I. Common response categories for selected questions

<table>
<thead>
<tr>
<th>Name</th>
<th>n (Q1)</th>
<th>n (Q2)</th>
<th>n (Q3)</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faster</td>
<td>12</td>
<td>11</td>
<td>8</td>
<td>&quot;The particles move way faster&quot;</td>
</tr>
<tr>
<td>Movement</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>&quot;The particles start vibrating more&quot;</td>
</tr>
<tr>
<td>Energy</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>&quot;There is more energy being put into it&quot;</td>
</tr>
<tr>
<td>Frequency</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>&quot;A higher pitch has more waves&quot;</td>
</tr>
<tr>
<td>Amplitude</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>&quot;The waves would be a lot bigger&quot;</td>
</tr>
<tr>
<td>Density</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>&quot;The dots would be more condensed&quot;</td>
</tr>
<tr>
<td>More Air</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>&quot;There would have to be more air particles&quot;</td>
</tr>
</tbody>
</table>

V. DISCUSSION

The unmatched prevalence of the faster responses across all three questions necessitates an immediate recognition: the notion that air particles move faster in a louder or higher-pitched sound wave is not true! Sound waves traveling through air are alternating waves of high and low pressure. As a speaker moves forward, some number of air particles that would not have contacted the speaker had it been stationary are instead reflected back into the same region as other unaffected particles, causing an increase in density. Then, as the speaker retreats, some number of air particles that would have reflected back if the speaker been stationary are allowed to travel slightly further “into” the space previously occupied by the speaker, causing a decrease in density. When sound is louder, the speaker moves with a larger amplitude, increasing the number of particles affected and causing the magnitude of the density fluctuations to increase. When sound is higher pitched, the rate at which the speaker oscillates increases. This creates more alternating density zones in a given period of time, which also causes each zone to be thinner.

Meanwhile, the speed of the air particles is dictated by their kinetic energy, which is measured via temperature; if the temperature of the air remains constant, the average speed of the particles will also be constant. Air particles at standard temperatures and pressures move at roughly the speed of sound, between 300 and 400 meters per second. Determining the speed of a typical speaker is challenging as it will vary based on the size of the diaphragm and the pitch and volume being produced; However, conservative estimates put most speakers’ speeds at roughly an order of magnitude or more slower than the speed of sound [23]. As such, the speed of the air molecules which happen to ricochet off of a speaker’s diaphragm will be relatively unchanged. Furthermore, the speaker would contact air particles just as often on its forward and backward strokes, resulting in a near-zero net impact on the average speed of the particles on the whole [24].

Digital simulations which attempt to model sound waves in air are faced with several pragmatic impossibilities: not only would a simulation which depicts air particles at a somewhat-accurate speed and randomness be computationally taxing, the volatility of the display would likely render it incomprehensible. As is the case with many physics concepts, simplifications are necessary when designing usable sound wave simulations. As seen in the examples provided [13, 18–20], the most common means for addressing these concerns is to program air particles which are semi-stationary and oscillate in tandem with their neighbors, creating the illusion that they are connected in a loosely-coupled lattice. This accommodation successfully avoids the previously mentioned pragmatic concerns, but inadvertently makes the simulation more susceptible to misinterpretation.

We believe that, when confronted with the sensemaking challenge in the sound wave module, the students were generating mental models for sound waves which took the simulated air particles to be literally analogous to real-world air particles. As the students observed the digital simulations, the motion of the air particles would stand out as an easily perceptible feature and would cue familiar resources regarding classical motion. According to KiP, these resources would then be the primary driving force behind the students’ reasoning as they navigated this novel context. When Q1 asked the students to predict the effect of increasing the speaker’s volume, it would not be surprising to see the speed of the particles play such a prominent role in the students’ answers, given how attention-grabbing the motion was in comparison to other visible features in the simulation.

When answering Q1, the students were theorizing about increasing the speaker’s volume but could not confirm their predictions in the moment as the digital simulations were not interactive. However, the NetTango coding environment let students alter the parameters of the environment and observe the consequences immediately. Thus, for Q2 and Q3, the students could observe the simulation’s actual response to the varying volume and pitch. As the interactive simulation modeled air particles as semi-stationary and loosely coupled, increasing the volume and pitch did, in fact, cause the programmed particles to move at a higher speed. This confirmation would reinforce the previously activated resources as reliable and trustworthy, solidifying their connection within the students’ knowledge network.

We refer to the normative model for sound waves in air as the “Gas” model and the proposed students’ model as the “Lattice” model, visualized in Figure 2. The Gas model describes air particles as moving independently and randomly, with sound waves being formed of alternating zones of high
and low density. In this model, a louder sound wave would have a larger disparity between the high and low density regions. In comparison, the Lattice model envisions air particles as being connected to their neighbors and subject to restorative forces that allow them to move about or oscillate while always returning to an equilibrium position. In the Lattice model, a louder sound wave would displace the particles further from their equilibrium position.

In alignment with our KIP framework, we wish to emphasize the utility of the students’ Lattice model even if it may be non-normative within the context of air particles. While semi-stationary particles may be inaccurate for gases and other fluids, the Lattice model is remarkably accurate for solids. Going further, simulations which depict coupled particles can still help students visualize critical aspects of the microscopic mechanisms for sound in air particles, namely the lateral propagation of high-density regions moving independently of individual particles. As such, we believe that the utilities and relevancies of lattice-style simulations more than justify their place in physics instruction; their inclusion just needs to be accompanied with proper framing to help students navigate the challenging nuances. Here, we offer a few suggestions for instructors using digital simulations for sound wave instruction in physics classrooms.

First, instructors should introduce the simulation with proper context. Depending on the placement of the sound instruction within the larger curriculum, students may have already been introduced to a particle-based microscopic view of the various states of matter. If this is the case, the instructor should have students recall the chaotic, random motion of particles seen in gases and compare it to the motion of the particles in the simulation. Instructors can directly address the simplifications made when modeling gas particles and ask students to predict how these simplifications may affect the validity of conclusions drawn based on the simulation.

Second, instructors should help students attend to the most relevant aspect of these simulations when discussing gases in particular. According to KIP, students’ sensemaking is widely shaped by the characteristics and properties of a phenomenon which receive the greatest attention or trigger recollection of familiar patterns [16]. This means that instructors can drastically influence the direction of students’ thought process by priming relevant connections or placing focus on important facets of the phenomenon. Instructors should foreground the density fluctuations over the lateral motion of the individual particles. Students can explore how the oscillating motion of a speaker would create such regions, and then be challenged to consider how their proposed mechanism compares with both the lattice-style simulation and the random motion of real-world gas particles. This conceptualization focusing on propagating density waves can then be connected to and supplemented by other classic demos or thought experiments, such as the suspended particle or candle flame vibrating at the mouth of a speaker [7, 10]; instead of imagining the vibrations as being in sync with oscillating, semi-stationary air particles, the motion can be explained by passing density waves creating pressure gradients in alternating directions.

Third, instructors should promote the Lattice model for sound waves within solids, where particles are connected to their neighbors via intermolecular forces; such a model provides valuable insights when discussing topics such as the longitudinal waves traveling across the string in the “lover’s telephone” [9]. The Lattice model can also be given practical importance by connecting it to the study of seismic waves: primary waves (P-waves) from earthquakes are longitudinal and are accurately represented by the model, while secondary waves (S-waves) are transverse. Interestingly, this modeling may help explain why P-waves can travel through both the solid and liquid regions in earth’s core while S-waves cannot: students would need to consider how the coupled motion within the solid regions would be converted into density fluctuations in the liquid regions (essentially, considering the solid material at the boundary to act like an oscillating speaker).

With these adjustments, instructors can make students explicitly aware of the differences between the simulations and reality, indicate which characteristics of the simulation are relevant to fluid-based contexts, and validate the Lattice model seen in the simulation for solid-based contexts. We believe these accommodations can allow instructors to make the most of digital simulations in their classrooms while making use of productive student ideas and validating students’ conceptualizations.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (1842375, 1941524).

[23] Human hearing is most sensitive to sounds of a few thousand Hertz. In order for a speaker oscillating at those frequencies to move at a speed comparable to that of air molecules, the amplitude of the oscillations would need to be at least several centimeters and perhaps more than a dozen. As most speakers will have diaphragms which oscillate on the order of millimeters, their speed must be significantly less than the speed at which air particles travel.
[24] In an ideally isolated system, any energy supplied to a speaker would eventually accumulate as kinetic energy within the air particles and raise their temperature. However, for any real-world system, the energy input of a speaker would be infinitesimal compared to the total kinetic energy of the system or the ambient heat flowing into or out of the system.
Modeling confusion in collaborative learning

Peter Bagdovitz
Western Washington University, Bellingham, WA, 98225, USA

Jayson M. Nissen
Nissen Education Research and Design, Slidell, LA, 70458, USA

When students experience confusion, resolving that confusion can lead to deeper understanding and engagement. Persistent unresolved confusion, however, can lead students to frustration and disengagement. Our research explores confusion and other emotions associated with learning as students work through elicit confront resolve (ECR) activities in Next Generation Physical Science and Everyday Thinking (Next Gen PET) physics courses for pre-service elementary teachers. We used the experience sampling method (ESM) to measure students' subjective experiences during seven particularly confusing activities. The ESM asked about confusion, self-efficacy, engagement, and stress, which we chose to align with existing models of confusion in learning. After some revision, our model fit the data well using confirmatory factor analysis. The only activity that required students' consistent use of mathematics produced the highest levels of confusion and stress for students. This relationship with mathematics content indicates the Next Gen PET courses could better support pre-service elementary teachers developing fluency and comfort with mathematics.
I. INTRODUCTION

While students and instructors often perceive confusion as a stumbling block in the learning process, confusion acts as a catalyst for growth and development [1]. As students navigate through the complexities of physics courses, they must confront confusion often. When students resolve their confusion, they both learn and gain a sense of achievement, but being stuck and confused can lead to stress and disengagement. To support students learning through complex tasks, research-based materials often have students work in pairs or small groups. Student groups’ ability to manage confusion and the attendant emotions that may accompany it varies. Some groups maintain a productive focus on collaborative physics learning, while confusion may derail other groups and lead to frustrated and disengaged students.

Physics educators often use collaborative learning strategies [2, 3] in the design and implementation of research-based, student-centered instruction. Student-centered instruction in physics, such as Tutorials in Introductory Physics [4] and Next Generation Physical Science and Everyday Thinking (Next Gen PET) [5], often uses an elicit confront resolve (ECR) framework. In an ECR framework, classroom activities present students with scenarios known to produce confusion and students work together to resolve their confusion. Curricular materials typically scaffold this resolution; however, they focus on conceptual understanding with little explicit support for students in managing social interactions or emotions.

To investigate these confusing experiences, we used the D’Mello et al. [1] model of affect shown in Fig 1. We adopt D’Mello’s description of confusion as an epistemic or a knowledge emotion. Confusion arises due to information-based assessments of how incoming information aligns with existing knowledge structures, as well as inconsistencies within the incoming information [1]. This model led us to measure stress, self-efficacy, engagement, and confusion during these activities. D’Mello and colleagues’ model applies to students engaged in learning activities who reach an impasse, such as in ECR-based activities. When students detect an impasse they may become confused. If they become confused they can then resolve their confusion through problem-solving and teamwork and continue to engage in the activity. If the activity was very confusing, resolving the confusion may create a sense of accomplishment and self-efficacy. Failure to resolve confusion can lead to stress and eventually disengagement. Identifying these emotions can identify topics or activities that elicit confusion and inform the extent to which the curricular materials support students in learning from their confusion or fail to support the students.

II. RESEARCH QUESTIONS

Our research was driven by two questions.
1. How well does the proposed factor structure of the
   items on the Experience Sampling Method (ESM) questionnaire align with the collected data?
2. What activities and other emotions are associated with higher levels of confusion?

The relationships between activities, confusion, and other emotions can allow us to identify activities where students failure to resolve their confusion leads to stress and disengagement. Identifying these confusing activities will assist our broader project on confusion and inform changes to the Next Gen PET curriculum to improve student cognitive and affective outcomes and collaboration.

III. METHODS

The research occurred in introductory physics courses for pre-service elementary educators. The courses used the Next Gen PET curriculum [5] and were taught at two primarily undergraduate institutions on the west coast of the United States. We collected data during 2021 and 2022 in nine courses taught by six different instructors. The summer 2021 courses were taught online and all other courses were taught in person.
The Next Gen PET curriculum prepares students to teach elementary-level science by covering a wide range of physics topics and focusing on students developing scientist-like views of these physics topics [6]. In the classes where we conducted the research, students worked in groups of 3-4. The students worked through activities that were part of the units for energy-based models (UEM), force-based models (UFM), and combination of forces (UCF). These units are divided into activities designed to cover one class period (e.g., UEM-A6 or UCF-A2).

We used the experience sampling method (ESM) [7-9] to measure students’ subjective experiences on four latent variables and one question on confusion. One-hundred and sixty students completed 650 ESM surveys. As a part of the ESM, the students completed brief surveys during activities that the course instructors and members of the research team identified as especially confusing for students. The research team marked the class activities with stickers to prompt the students to complete the surveys. At one institution, the stickers included a QR code that directed students to complete the survey online. At the other institution, the stickers prompted them to complete and turn in a paper survey that was inserted into the activity. The research used paper surveys because some students may not have had a device to access an online survey in class.

The ESM data collection was part of a larger systematic investigation of students’ confusion and socio-metacognition [10] in Next Gen PET courses. All students completed the ESM during class and a reflection prompt about the course outside of class. Researchers video recorded two to three groups of students in each section. Students from these recorded groups participated in interviews on their experiences. Other researchers in the project are analyzing these other data streams, but that data will not be discussed in this paper.

We compared four emotions across seven activities the instructors identified as especially confusing for students. These four emotions were stress, self-efficacy, confusion, and engagement. We also collected data on teamwork, but we did not include that measurement because student responses all had ceiling effects with little variation. To compare the activities, we used descriptive statistics, data visualizations, and the Wilcoxon rank sum test. These methods allowed us to identify the extent to which students experienced confusion in each activity and to identify an activity that elicited more confusion than the other activities. We then compared how stress, engagement, and self-efficacy compared in the activity where students reported high levels of confusion to the other activities. The Wilcoxon rank sum test is a non-parametric alternative to the paired t-test for independent samples. Researchers can use an equivalent to correlation, $r$, as an effect size [11]. The data visualizations overlaid a plot of the individual data points on top of the box plots.

A. CFA

We conducted confirmatory factor analysis (CFA) [12] to test the construct validity of our ESM survey. CFA is a type of structural equation modeling that tests the correlations in a model proposed by the researcher. We used a three-step CFA process to come to a model that fit the data well. In the three steps we (i) created an initial model based on theory and prior research, (ii) ran CFA and generated fit statistics using the lavaan [13] package, and (iii) used the modindices command to improve the fit. We repeated these last two steps until the factor loadings in the model were $> 0.6$ and the fit indices passed the cutoffs discussed in the next paragraph. Hair et al. [14] proposes factor loadings of 0.5, explaining 25% of the variance in the item, as an absolute minimum and 0.7 as a preferred minimum. 50% explained variance.

We used the root-mean-square error of approximation (RMSEA), comparative fit index (CFI), and Tucker-Lewis index (TLI) to inform how well the factor structure fits the data. RMSEA is an absolute fit index that addresses parsimony in the model by accounting for the degrees of freedom in the model. High RMSEA indicates an over-constrained model with too few degrees of freedom. RMSEA has a cutoff of 0.06 and below [15]. CFI and TLI are relative fit indices. They compare the test model to a baseline model with all covariances set to zero and all variances freely estimated. In other words, CFI and TLI compare the test model to a very poor model, so they look very good for models that fit the data reasonably well. Both indices range from 0 to 1, where 1 is the best possible fit and 0 indicates no fit. We used a cutoff of CFI and TLI $> 0.95$ [15]. A combination of absolute (RMSEA) and relative (CFI and TLI) fit indices allowed us to increase the fit of the model without over complicating the model. This balance leads to a model that is simple and fits well.

IV. RESULTS

Our research was driven by two questions. How well does the proposed factor structure of the items on the ESM align with the collected data, and what activities and other emotions are associated with higher levels of confusion? To answer the first question, we discuss the CFA and path diagram. We then use the latent factors for confusion, stress, self-efficacy, and engagement to investigate students’ confusion.

The iterative process for improving our CFA model produced a factor structure (Fig. 2) that fit the ESM data well. The standardized factor loadings for each item exceeded our cutoff of $> 0.6$. Fit indices also passed our cutoffs of RMSEA $< 0.06$, CFI and TLI $> 0.95$.

The initial model differed from the model shown in 1 in that the initial model included a question on how in control students feel under the latent factor for self-efficacy and a question on how much students wanted to do the activity under the latent factor for engagement. The factor loadings
FIG. 2: Diagram of the final factor structure. The boxes are questions on the ESM, the ovals are latent factors, and the factor loadings are the arrows linking them.

FIG. 3: Confusion, stress, self-efficacy and engagement across all seven activities. The scale ranged from 1, not at all, to 5, extremely. The lighter points in gold represent the individual responses from students with a slight ‘jitter’ to space them apart. The solid black dots are outlier responses for the boxplot. The activities are curricular materials meant to be covered in one course period and following the naming convention in the curriculum.

for these two items were below the cutoff of 0.6. We first removed the want to do question from the model and then removed the control question.

The following results revealed the activities and emotions that are associated with higher levels of confusion. Figure 3 presents students’ reported confusion, stress, self-efficacy, and engagement in seven activities. In most activities, most students experienced low levels of confusion. Median scores for confusion were between 1, not at all, and 2, somewhat, for all activities besides the sixth unit of the energy section of the class, UEM-A6.

The confusion in UEM-A6 stands out from the other activities. In UEM-A6 students experienced a median confusion of 3; the typical student participating in this activity was moderately confused. The interquartile range for UEM-A6 does not extend to 1, so few students worked through the activity without becoming confused at all. A large cluster of students, shown in gold in Fig. 3, responded as experiencing extreme confusion in UEM-A6, more than in all other activities combined. A Wilcoxon signed rank test of confusion in UEM-A6 versus confusion in all other activities showed that this difference was statistically significant, \( p < 0.001 \), with an effect size of \( r = 0.27 \), with confidence intervals of \([0.19, 0.35]\).

Comparing the stress, self-efficacy, and engagement in UEM-A6 to the other activities reveals some patterns consistent with the D’mello model and some inconsistencies. Stress was higher in UEM-A6 with a median of 2 compared to 1 or 1.3 in other activities, and more students reported very or ex-
treme stress in UEM-A6 than in all other activities; $p < 0.01$, $r = 0.31$ [0.23, 0.38]. Self-efficacy was lower than in most other activities. This difference was not as large as for confusion or stress. The difference in self-efficacy compared to all other activities was small $p = 0.06$, $r = 0.07$ [0.00, 0.15]. Students reported similar engagement in UEM-A6 to the other activities, $p = 0.33$, $r = 0.04$ [0.00, 0.11], and they did not report the disengagement the D’mello model indicates high levels of confusion can lead to, see Fig. 1.

V. DISCUSSION

We studied seven activities instructors identified as particularly confusing, but as Fig. 3 shows students tended to report ‘slightly’ to ‘not at all’ confused in all activities besides UEM-A6. This contrast between students reporting little to no confusion and instructors identifying these activities as particularly confusing may indicate a low level of metacognition for the students, or it could indicate that instructors identified activities where some students were very or extremely confused. This difference between instructors and students could, however, have resulted from a difference between what the Next Gen PET materials expect students to do and what instructors with PhDs in physics think of as a full understanding of these topics. Further exploring students’ metacognition in these activities can inform the extent to which students engage in metacognition and the ways instructors and curricula can scaffold students’ metacognition.

Descriptive statistics indicated the ECR activities aligned with many aspects of the D’Mello et al. [1] model of affect. In the UEM-A6 activity, where students reported the highest levels of confusion, students also reported higher stress but only slightly lower self-efficacy. The elevated stress, especially in activity UEM-A6, indicates some students did not resolve the confusion elicited by the activity and likely failed to achieve a sense of accomplishment.

In UEM-A6, students did calculations involving fractions and percentages in each step of the activity. Students could not circumvent the mathematical sections. Only UEM-A6 had significant mathematics required in the activity.

We cannot conclude how engagement correlates with confusion within this study. Because the ESM only captured a snapshot of students’ experiences, the data cannot clarify if they stayed engaged over the whole activity or disengaged at a later point. To address this limitation of the ESM, we are also using classroom videos, interviews, and journal reflections to study the confusion, stress, and engagement students experienced in Next-Gen PET with a focus on UEM-A6 as the most confusing activity. For example, one student who found UEM-A6 very confusing and stressful further engaged with the material by seeking help from instructors outside of class.

VI. CONCLUSION

The Next Gen PET curriculum seldom requires students to do mathematics. UEM-A6, however, required students to do mathematics throughout the activity. The higher levels of stress and confusion that students experienced in UEM-A6 than the other confusing activities indicates that the mathematics content and students’ mathematics anxiety may have been the primary source of confusion and stress in this activity. Next Gen PET courses primarily serve pre-service elementary school teachers. Elementary school teachers tend to have higher levels of mathematics anxiety than any other professions [16], and they experienced their highest levels of mathematics anxiety in university [17]. Next Gen PET instructors could integrate more mathematics into their activities to support students in developing confidence and fluency for mathematics. Supporting pre-service teachers in developing this fluency and confidence would likely increase the mathematics learning of their future students [18].

The D’Mello model [1] focuses on several emotions students experience while learning and experiencing confusion, but none of these affective emotions address students’ predispositions towards a course or content area, such as mathematics anxiety. As shown in Fig. 1, the D’Mello model posits the student begins in an equilibrium state characterized by a feeling of engagement, but a student experiencing mathematics anxiety may begin an activity with similar engagement to their peers but with higher stress levels than their peers. They may even feel stuck or stressed once they realize the activity requires them to do mathematics. The D’Mello model implies that students experiencing this elevated stress may be more likely to perceive themselves as stuck and shift to disengagement than they would in other activities or than their less anxious peers. Students with higher levels of mathematics anxiety likely need more support and scaffolding from their peers and instructors and the curriculum to resolve their confusion, gain a sense of achievement and self-efficacy, and maintain their engagement in the activity.

VII. ACKNOWLEDGEMENTS

This work was made possible by funding from the National Science Foundation (Grant No. 1928596). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We also thank Thanh Lê, Andrew Boudreaux, and Carolina Alvarado for their work on the project.


“Science happens between people”: teachers’ perspectives in a physics RET program

Roshmi Bano (she/her)
Department of Chemistry, University of Illinois at Chicago, 845 W Taylor St, Chicago IL 60607

Minjung Ryu (she/her)
Department of Chemistry, University of Illinois at Chicago, 845 W Taylor St, Chicago IL 60607
Learning Sciences Research Institute, University of Illinois at Chicago, 1240 W Harrison St, Chicago IL 60607

RET summer programs are envisioned as spaces that can help bridge the gap between classroom STEM teaching and STEM research. Much research has focused on the impact of RETs on teachers’ beliefs about inquiry based learning. RET programs provide a unique setting where teachers, as short term visitors in the research lab space, encounter and engage with ideas about the practice of science itself. In this work in progress, we present the narratives constructed by two physics teachers about the doing of science and the doers of science in a quantum science summer RET program. Using a qualitative approach, we analyze teachers’ interviews over the summer and show how these teachers constructed their identities with regard to the doing of science. We demonstrate how both teachers, by the virtue of trying on these new identities, repositioned themselves as doers of science and challenged dominant ideas about the doing of science.
I. INTRODUCTION

Research Experience for Teachers (RET) programs aim to engage K-12 science teachers in research practices of scientists in university-based laboratory settings and bridge STEM research and classroom practice. Scholarship among science education researchers has focused on understanding how teachers experience STEM research in these programs and how their participation affects their beliefs about classroom science teaching [1–3]. Some scholars have utilized teacher identity as a construct to understand the changes that science teachers go through as they participate in RET programs [4,5]. Teacher identity is a powerful construct to situate and describe trajectories of teacher learning [6]. Thus, in this paper, we use teacher identity to show our preliminary results on how two high school physics teachers in a quantum physics RET program view the doing of science, position themselves as doers of science and try on new identities in this new space outside of their classrooms.

II. THEORETICAL BACKGROUND

In this section, we briefly review the extant literature on research on teacher experience in RETs. Next, we lay out the use of teacher identity as a lens to understand teacher change, and a few important studies in the RET context that used this lens. We then delineate some current perspectives on the meaning of doing science. Finally, we state how we wish to use the identity lens to explore how physics teachers in an RET program position themselves as doers of science in a research laboratory context.

A. Research on teachers’ RET experiences

Since the widespread implementation of RETs as an important opportunity for in-service teachers’ professional learning, much research has focused on evaluating the impact of these programs. Particular attention has been paid to how these programs affect teachers’ views regarding inquiry-based science instruction. Work by Hughes and colleagues showed that participation in a summer RET program improved primary and secondary teachers’ understanding of the scientific inquiry process [1]. In another instance, a summer RET program followed by a yearlong professional learning community shifted teachers’ beliefs about science classroom practice from a teacher-centered approach to a guided or structured inquiry approach [3].

B. Teacher experience in RETs through the identity lens

Many scholars on teacher identity have centered the conceptualization of this construct around Gee’s framework [7] of identity as “the kind of person one is seeking to be and enact in the here and now”. Building on this, Beijaard and colleagues [8], in a literature review of teacher identity, identify that an essential feature of identity is that it is “an ongoing process of interpretation and reinterpretation of experiences” and that “it implies both person and context”. We align our analysis in this paper with Beijaard and colleagues’ concept of teacher identity and focus on how teachers interpret and reinterpret their experiences in the context of an RET summer program to construct their identities as science practitioners.

We highlight important findings from previous work that has pioneered the use of the teacher identity construct in examining teacher experiences in RET programs. Varelas and colleagues [4] utilize teacher identity to understand how teachers make sense of their practices as scientists and science teachers and how they navigate the emergent tensions between their identities. In more recent work, Davidson and Hughes [5] investigate factors that influence teachers’ level of belonging in the “community of science practice” and find that both program design factors such as research group structure, as well as personal factors such as teachers’ sense of expertise in doing the research work are important.

C. Ideas about doing science

In understanding science teachers’ identity construction, it is important to examine existing ideas about science and science teaching. Holton states, “[there] are two very different activities, both denoted by the same word, ‘science’: the first level of meaning refers to private science, the science-in-the-making... the second level of meaning refers to the public science, science-as-an-institution, textbook science, our inherited world of clear concepts and disciplined formulations” [9]. Holton’s framing makes a distinction between ideas of science held by the general public as established and institutionalized versus a more tentative and constructivist idea of science held by practitioners such as scientists and researchers. Furthermore, a recent study by Warren and others [10] calls out this institutionalized science attributing it to the invisibilization of the “dynamic nature of disciplinary activity where disciplines are flattened and engaged with as if they are static, known and finalized domains”. In the context of science, they emphasize that this “settled” disciplinary knowledge is shaped by normative ideas about who is allowed to create scientific knowledge. In this paper, we refer to these individuals as doers of science.

It has been shown that discourses about science teaching (for example, traditional science teaching Discourse versus teaching science through inquiry Discourse) can influence what identity options are available to and are taken up by teachers [11]. We therefore conjecture that dominant ideas about doing science (e.g., science-in-the-making versus science-as-institution) can also influence the identities
available to teachers as be doers of science in the research lab. Thus, in our preliminary analysis, we aim to understand how two physics teachers in an RET program construct their identities with regard to doing science, by positioning themselves and others as doers of science.

III. METHODS

A. RET program context

Our study examines two physics teachers participating in a paid six week summer RET program focused on quantum science, primarily based in an interdisciplinary science research center of a private research university. Participating teachers applied for the summer program in the preceding spring. Each participant was referred to as a teacher researcher, hosted by a research lab led by a faculty member and matched with a research lab member (graduate student, postdoctoral researcher or staff scientist) who acted as their research facilitator. During the first week of the program, teacher researchers spent most of their time getting oriented towards working in the labs and establishing relationships and planning projects with their research facilitators. In addition, they participated in “quantum bootcamp” where quantum science concepts were reviewed by a participating faculty member in a lecture format. Starting in the second week of the program, teacher researchers spent most of their time in research labs, with one morning per week dedicated to a curriculum development session, another morning per week dedicated to curriculum group work and one afternoon per week at a content instruction session delivered by a faculty member.

B. Participants and data collection

The main data source for this study was semi-structured interviews conducted with all five teacher participants at the beginning and end of the six week summer program. All interviews lasted around one hour and consisted of questions about teachers’ thoughts and feelings about the RET programming, their interactions with other teacher researchers, research facilitators and faculty members, and their experience navigating the research lab space. We analyzed interviews from two teacher participants in this preliminary stage of our work. Both Mr. Martin and Ms. Adam held degrees in physics and were physics teachers. Mr. Martin had 2 years of teaching experience, while Ms. Adam had 21 years of teaching experience.

C. Data analysis

All interviews were transcribed verbatim. We drew on the principles of phenomenological analysis, and analyzed the interview transcript data to make meaning of the teacher researchers’ experiences in the RET program. The unit of analysis was the teachers researchers’ reflection on a specific event or activity in the RET program such as doing an experiment in the research lab or having discussions with the research facilitator. The first author composed multiple analytic memos on these individual descriptions to understand how these reflected participants’ underlying views on the doing of science and the doers of science. Finally, memos integrating thoughts from the two interviews from each teacher were written to understand how and why teacher researchers’ views evolved through the course of the RET program.

IV. FINDINGS

A. Mr. Martin: owning the teacher identity

Mr. Martin was excited for his research lab experience – the prospect of having hands-on experience with in the laboratory made him feel “over the moon.” In Mr. Martin’s first interview in the first week of the summer program, when asked about working together with his research facilitator, he expressed an interest in being as helpful as possible. He also mentioned his thinking on how he could be helpful in the research lab, saying:

“And they are all very keen on me coming in as a science teacher, not so much as a scientist, which I think is the entire point of the program. I need to sometimes switch that perspective in my brain, because sometimes I’m looking at things from the perspective of a researcher, not from the perspective of a teacher.”

First, Mr. Martin distinguished two different kinds of science practitioners he envisioned being in the research lab space – a science teacher and a research scientist. Interestingly, he also spoke of switching between these two practices, suggesting that he thought of them as mutually exclusive. Second, he got the sense in the initial days of the program that his research lab valued the science teaching skills he brought. We focus on these two threads in Mr. Martin’s first interview.

We first explore Mr. Martin’s ideas about the practice of science, and how his thinking was exhibited during his research experience in the summer program. In the second interview at the end of the summer program, Mr. Martin recalled that he was very curious to learn “the process of being a researcher.” In Mr. Martin’s account of his research activities, he described reading research articles, learning experimental procedures and fabricating a device that was used to obtain research data. Thus, ostensibly Mr. Martin successfully carried out all the research activities planned for him, and yet when he spoke of his feelings when swiping his ID to walk into the research lab, he said, “I felt very cool. […] And I was like, “Y’all don’t even know. I have no idea what I’m doing.” What drew our attention is the contradiction in Mr. Martin’s description – despite the research work that Mr. Martin engaged in, it seemed hard
for Mr. Martin to feel like he belonged to the research space, perhaps like other members of the research lab, like graduate students. Another piece of data hinted at how Mr. Martin distinguished his science practice and his research facilitator’s science practice:

“So there is definitely a healthy mix between my experience as a teacher and providing, and being the bridge to content and having that ability to translate that content to high school level. And then [research facilitator]’s expertise in knowing true physics, if that makes any sense, a true understanding of quantum concepts.”

Here, Mr. Martin was speaking about collaborating with his research facilitator to develop a curriculum module based on his summer research experience. We note that he referred to his research facilitator’s expertise as “knowing true physics.” By positioning his research facilitator, a member of the research lab, as holding the “true” knowledge of physics, Mr. Martin drew on a dominant idea of normative science practice where only individuals who are engaged in science research in a laboratory at the graduate level and beyond are seen as truly doing science. Perhaps, this is why, in Mr. Martin’s view, he felt like he didn’t know what he was doing, because even though he was in the research lab space, he was still a science teacher and not a science researcher like his facilitator.

However, positively positioning himself as a science teacher in the research space is also what allowed Mr. Martin to challenge this dominant idea of normative science practice. Mr. Martin shared the experience that was central to this:

“In the beginning I came in and I was very anxious because I have not breached these higher tier scientific subjects in so long. I was thinking, “Well, I'm going to be coming in. And there's going to be an 18 year old who will blow me out of the water with their mathematical and physical physics expertise.” And then I come in and here I have some 18 year olds calling me, sir, and asking me about my expertise in teaching the subject. Because they are very clear in the idea that science happens between people. And if you can't explain what you're working on, you're not going to get anywhere. Nobody does a full project by themselves. They have to be able to correctly communicate what they're working on to people of various degrees of content knowledge. [...] But the part where they were actually interested in scientific communication was the part that made me realize like, oh wait, I have knowledge and skills that can be beneficial to the people in the lab.”

For Mr. Martin, members of the research lab respecting his expertise in communicating science as a teacher was crucial to recognizing he had important skills to contribute to the research space. In fact, Mr. Martin told us that he saw his role in the research lab as someone that lab researchers could practice explaining scientific concepts to. The experience of being positioned as a science teaching expert by research lab members led to Mr. Martin thinking about doing science in a different way, where he saw conversations between people about scientific work as integral to the process. Thus, while Mr. Martin retained the distinction between science teacher and researcher, being in the research space with a strong teacher identity allowed him to challenge this binary and create a legitimate space of membership for himself in the research lab.

B. Ms. Adam: embracing a learner identity

Like Mr. Martin, Ms. Adam was excited to be in the research lab space. In her first interview at the beginning of the program, she told us that she felt privileged to be experiencing a “state of the art” science laboratory. At the same time, she was concerned whether she would be able to understand the quantum science content well and fit into the research lab. Ms. Adam’s perception of the research lab members and in particular, the professor was illuminating:

“I felt the professor was going so fast, but it was for me so fast. He actually was really, I think this is the best that he can bring it to the public, and that's part of the communication. He's too much knowledgeable, and he is around so many smart students.”

Ms. Adam was reflecting on the professor’s teaching in the quantum bootcamp. She positioned her research professor as “too knowledgeable”. She also positioned the other student members of the professor’s research lab as “smart” and in another excerpt, as “very, very, very advanced”. In doing so, Ms. Adam constructed an idea of the research scientist, an individual capable of creating scientific knowledge, as a person with extraordinary, and even unattainable intelligence.

Very early on in the summer program, however, Ms. Adam had an interesting experience in the lab space that provided her a different perspective about who can be a scientist. Ms. Adam was assigned a desk in her host research lab and noticed the graduate student researchers’ desks. She wanted to take pictures of these desks and share them with her students:

“And the clutter of the notes, and the coffee. It was like this is what the people have, so they can't see, relate. "Oh, a student is a student. Whether you are a grad, they have tests to do, they have activities. They have a nice thing. I think one had some of the KPOP poster or somebody. It was like, oh, those are normal. [...] And again, a lot of them that in the younger generation, so they can connect with my students. [...] Normalize them as a humans, and
just like. Someone is in the lab doesn’t mean that they have just to be... They may be nerd in their specialty and stuff, but that doesn’t mean they don’t have life, and they don’t have interest, and things.”

The everyday objects on the graduate researchers’ desks, such as coffee, clutter, and posters, humanized the unimaginably intelligent scientists for Ms. Adam. She describes thinking in this moment “oh, those are normal.” In the research lab space, Ms. Adam found a point of relatability with the graduate student researchers in that she saw them as just students and youths and, thus, similar to her own students.

Hearing Ms. Adam’s account of her summer of research in the second interview, it became clear that this early encounter was important to her experience. Ms. Adam reflected on her general experience of being in the research lab saying, “It’s put me on the feet of my students.” Imagining what her students might feel like in the research space and what they might like to know and see about it may have been central in feeling like she was in her students’ shoes. Additionally, Ms. Adam also told us that due to curriculum development work in the program, she was often thinking about how to share the content she was learning with her students. Such student-centered thinking, combined with being positioned in the lab as a research apprentice of sorts may have encouraged Ms. Adam to embrace and enact her identity as a learner in the research lab. Further conversation with Ms. Adam revealed her actions that crystallized into this identity:

“Like when she [research facilitator] will give me a homework, read these articles and come with the three questions. And I will go and say, okay, in the article, in this paragraph, I saw this. What’s that mean? So I felt it’s like a homework. So I kind of put myself as a student. I need to learn these materials. I need to apply these activities. I need to participate with this, whatever she will tell.”

Ms. Adam viewed the research lab as a flipped classroom, and herself as a student. She immersed herself into research activities facilitated for her and actively reached out to her facilitator with her questions. She even spent time outside of the research lab hours doing “homework” and reading research articles. At the end of the program, Ms. Adam felt settled into the research experience and told us that she saw it as a place where she could focus on her own learning. For her, working in the research lab became a routine where “everyday was something fun”. Thus, by stepping into the identity of an enthusiastic learner, Ms. Adam established her membership in the research lab space.

V. CONCLUSIONS

Our preliminary analysis demonstrated how two physics teachers constructed two different identities (teacher and learner) while navigating the summer RET program. These two identities, and more identities that we could not focus on, served as resources for them to reposition themselves as doers of science and legitimate members of the research labs in various ways. In doing so, they also challenged the notions of institutionalized science and reframed science as a more humanized enterprise that requires many different kinds of skills and centers human interactions.

In this study, the identity construct allowed us to notice the important learning teachers accomplished and the moments in which normative notions of science were breached. However, with the scope of our data, we cannot say whether this repositioning of science and doers of science was stabilized or not. Future studies should focus on how RET programs can be designed to allow and encourage this kind of repositioning and how the repositioning observed during the summer program impacts teachers’ classroom practices.

Our study provides useful implications for the design of RET programs. First, we note that the social context of the laboratory were instrumental in teachers’ repositioning. For example, Mr. Martin description of his lab colleagues’ interest in learning from his teaching skills, or Ms. Adam’s note about how the physical contents of the lab space made her feel. In the implementation of RET programs, it is thus important to attend to teacher-researcher relationships, encourage asset-based thinking about teachers’ experiences and promote intentional inclusion of teachers in the social structures of research groups. Second, teachers’ positioning of scientists and research practices was central to developing a deeper understanding of how scientific knowledge is produced and how there may be multiple ways to be involved in this process. Thus, it is important to realize that RET programs provide a space not just for the transfer of content knowledge, but also knowledge about the processes of doing science. RET participants are carriers of this knowledge to their high school classrooms. Attending to this aspect of RET programs allows further dispersion of knowledge to students in science classrooms who may be future doers of science.


Acquisition of qualitative video data: methods and reflections in PER

Lauren Barth-Cohen
Department of Educational Psychology, University of Utah, 1721 Campus Center Dr. SAEC 3220, SLC, UT 84112
Department of Physics and Astronomy, University of Utah, 115 S. 1400 E. 201 JFB, SLC, UT 84112

Tamara Young
Department of Physics and Astronomy, University of Utah

Jason May
Department of Physics and Astronomy, West Virginia University

Adrian Adams
Department of Educational Psychology, University of Utah

Video Data is commonly collected in PER to allow for insights into how learning and teaching unfold over time. One might think the collection of video data is straightforward, but there are key decisions about gathering video that can profoundly impact the entire project. Here we take a microscope at the common practices of gathering video data in PER. Through two existing cases in the qualitative PER literature, we describe how and why PER scholars made those key decisions. We open the black box of research planning and decision-making when video data is being collected. By increasing transparency, we aim for the community to better understand the decisions made behind the scenes in the research process, which may strengthen other PER scholars’ future research endeavors.
I. INTRODUCTION

Video data is widely used in qualitative research in PER [e.g., 1–3]. Video is important for allowing insights into how learning and teaching unfold over time, which has provided the field with new knowledge about student behavior and reasoning not accessible through audio or student artifacts. That is, through video analysis PER has gained important insights into social dynamics role in the learning and teaching of Physics, through analyses of gestures, body position, and facial expressions (e.g., gesture analysis, posture, social dynamics, positionality, etc.) [e.g., 4, 5]. One might think the collection of video data is straightforward. But, in a few weeks or months, one can amass a huge corpus of video, which can be overwhelming given the practical need of making decisions during the data analysis process. When large amounts of video are collected without careful consideration of the research goals, it can lead to ambiguities about next steps and the subsequent analysis. Adjacent fields, such as the Learning Science and Science Education have written about video analysis [6–9], but there are fewer new video methods pieces and there are differences across learning environments and emphases that can have ramifications in the gathering of video data.

Here we argue that there are key decisions to be made before and during data collection that results in the collection of the optimal amount for a given research goal. Those decisions can have a profound impact on nature of the video data collected and the entire project. Through two cases of PER articles engaged in video analysis, this paper aims to describe how and why PER scholars made those key decisions so that the larger PER community can better understand the decisions made behind the scenes of the research process, which in turn may strengthen their future research endeavors. By synthesizing key ideas and decision points around the gathering video data, we aim to illuminate these issues for a PER audience that might be new to video data collection and/or interesting in learning from these adjacent fields.

In what follows, we use two different cases from the empirical PER literature to highlight how video data collection decisions are impacted by the research goals, including the theoretical framing and assumptions. We specifically focus on decisions related to camera placement, the amount of video equipment used, and the level of detail needed in the videos. These two cases allow for a comparison of how video is used differently across learning environment structures (entire classroom environments versus small group work), including the roles of lab classroom technology and the instructor in the gathering of video data. Beyond these differences, across both cases, we show how the video data collection process was tailored to the specifics of the research goals. Thereby, we shed light on the behind-the-scenes research process of gathering video data in PER. Finally, we conclude with a series of questions and decisions points that are meant to support PER Scholars in planning their data collection in light of their research goals.

II. CASE 1: GATHERING VIDEO DATA WITHIN A PANORAMIC CLASSROOM PERSPECTIVE

Our first case of how and why PER scholars made key decisions about video data collection comes from a qualitative analysis of a classroom of 9th-grade students participating in an embodied modeling activity known as Energy Theater [10]. Energy theater was developed to engage “learners with key conceptual issues in the learning of energy, including disambiguating matter flow and energy flow and theorizing mechanisms for energy transformation.” [11]. In this instance, the students were modeling the steady-state energy of the Earth in an Earth science class. The teacher introduced the activity by creating space in the classroom for the activity and describing the roles: Each student represents one unit of energy. Ropes are used to delineate objects in the system (e.g., sun, earth). Movement from one point to the next indicated movement of energy between objects in the system. During the activity, students engaged in rich conversations about the key objects in the system, while negotiating details about the wavelengths of light and movement of energy in the system. Sometimes many individuals were talking at once, and sometimes everyone was silent while one person shared an idea aloud. A body of prior research has documented the advantages of this activity for supporting conceptual engagement about conservation, storage, transfer, and flow of energy along with the disambiguation of matter and energy [11–15].

During the class, the 9th-grade students were introduced to Energy Theater, its rules, and assumptions, and then told to model the steady-state energy of the earth. The authors collected data from three 45-minute class periods, each of which had two science classrooms that were team-taught by two teachers. One class period was not studied because the class sizes were so small that nearly no discussion happened, and discussion is key for video data that can be analyzed with this theoretical perspective. For each class period, the students first enacted Energy Theater in their classrooms. Then the two classes came together and watched their peers’ Energy Theater enactment. Finally, there was a joint enactment of Energy Theater with all students from both classes.

The authors were interested in the conceptual learning that may have occurred during this activity. To capture this learning, the authors used a theoretical framework known as Coordination Class theory [16, 17]. Knowing that this theory had previously been used to capture learning in interview settings, the authors were interested in the affordances and limits of this theory in capturing classroom learning within Energy Theater. As some background, Coordination Class theory sits within the larger Knowledge in Pieces and Resource theory perspective [18, 19]. This theory is a model of a concept within a learner’s larger knowledge system. Specifically, using this theory, one is typically aiming to capture how learners’ understanding of a concept (e.g., force: acceleration–or in this instance, steady state) changes over time through the application of the concept to new contexts, for instance, across different problems or learning environments. Within
the theory, a potential learning difficulty is that of determining the same information from different contexts, given that likely different knowledge is being used in those contexts (e.g., homework problems vs. laboratory experiments—or in this instance, different models created in Energy Theater).

A. Video Data Collection Considerations

Given the theoretical framework and the specifics of Energy Theater, the video data collection plan was tailored to the project. The authors aimed to capture video data from as many students as possible and their teachers as they moved around the classroom, acting out the energy transfer and transformations across the objects in the scenario. It was important to collect information on the student’s embodied actions with respect to the objects in the scenario, for instance, their position concerning the ropes representing the sun and earth. Needed was also information about their gestures or other body details that were used to convey information about their type or form of energy. For instance, students waved their arms back and forth, tag peers, and hold up their hands to signal different forms of energy. This type of information about embodied actions is foundational to the Energy Theater given its importance in the student’s models of Energy flow, transfer, and transformations, and therefore key to our research goals and subsequent analysis. To capture this, the authors placed cameras on tripods at the room edges in each classroom. Another goal was to collect high-quality audio data of the students’ discussions, as discussions are key for a Coordination class theory analysis. To collect this data, external audio recorders were placed around the necks of the teachers, given the likely importance of capturing good audio from them. We also hung a few audio recorders from the ceiling to further increase the likelihood of strong audio data from different locations in the classroom. Finally, we had unique concerns related to the logistics of the space (simultaneously collecting data in two team taught classrooms) and the scheduling of the classes (spare time between classes). Therefore, we placed equipment in both classrooms to avoid moving equipment during a hectic class. Given the classes’ sequential nature and the school’s short passing time, we had to check batteries and memory cards during the very brief classroom transitions and change the batteries and memory cards as needed to ensure no missing data.

B. Results and Limitations Given the Video Data Collection

The results found that the students changed their models in specific ways that better aligned their understanding of the scientific concept with their newly modified model. For example, in one class they recognized that in their model, energy was leaving the earth and going to the sun, which was nonsensical and not what they intended to represent. They subsequently changed their model so that when the people (who represented energy) left the earth, they exited the model, and then reentered the model in the sun. In this case, it was a single student who pointed out the issue during a group conversation when everyone was listening. We were able to show how a single student’s contribution can dramatically affect the model and subsequent learning. Our ability to capture these findings was due to having the video of the entire class period, all the students’ movements around the classroom, and clear audio of the whole class discussions. We likely would have missed these crucial details about how and why they changed their models if we had collected only written notes or only audio data. Furthermore, we might have missed key details if we had used fewer cameras.

However, there were limitations of the data we collected and that had ramifications. We could not analyze each student’s individual learning over the entire class period. We also couldn’t follow every idea that surfaced during the class discussion, and some ideas were lost to the chaotic parallel discussions. We saw some embodied actions that we could not connect to verbal explanations. These limitations meant that the analysis focused on learning at the classroom level, not the individual level. But, importantly, learning at the classroom level through the coordination class theory lens was the crux of the contribution given that previously coordination class theory had been used to capture individual learning, often in interview settings.

III. CASE 2: GATHERING VIDEO DATA OF SMALL GROUP WORK IN AN INTRODUCTORY LAB COURSE

Our second case of how and why PER scholars made key decisions about video data collection comes from a qualitative analysis of four undergraduate students enacting group sensemaking in an Introductory Physics for Life Science (IPLS) major’s lab course [20]. The IPLS lab course in question had been redesigned to emphasize a Three-Dimensional learning (3DL) approach [21] and included structures so that student groups had the experimental agency to develop and carry out experiments that explore the physical properties of a biological phenomenon. For the article, we focused on a lab where the students were tasked to create an experiment to study Brownian motion to provide insight into how diffusion occurs. The students were provided with materials, such as synthetic microspheres suspended in fluid and microscopes. They also used an image-tracking software on a classroom computers that was stationed at their labs. This article sits within a body of research that is broadly focused on students’ reasoning processes in various intro Physics labs that have been reformed [e.g., 22–24], but less work has focused on the moment-by-moment reasoning processes in these labs. For this article, the data collection happened through in-person class observations of students’ small group work in the labs. The course mainly enrolled upper-division undergraduate students from life science disciplines (e.g., biology, kinesiology), and for most students, this was their first physics lab
course. During the semester, students worked in small groups of 3 or 4 individuals on multi-week investigations where they would develop research questions, experimental design plans, and hypotheses, and then investigate, develop a scientific argument, and present their argument to peers verbally and in written form. Importantly, the article’s first author served as TA for several of the observed sections, all of which used the lab classroom, and the fourth author served as the instructor of record for the course.

In this research, we were interested in students’ sensemaking in these labs. Specifically, the moment-by-moment details of their sensemaking process. In the analysis, we focused on a series of inconsistencies in their sensemaking, specifically what the inconsistencies are about and what moves students enacted to resolve them. We view sensemaking as a “dynamic process of building or revising an explanation to ‘figure something out’—to ascertain the mechanism underlying a phenomenon in order to resolve a gap or inconsistency in one’s understanding” [25]. At a high level, we focus on two distinct elements of the sensemaking process: the process of “figuring something out” through explanation construction, and the nature and recognition of the inconsistency being resolved in the sensemaking process. Within this theoretical framing, we had the following research question: What forms of inconsistencies are students in introductory physics lab courses sensemaking about, and what moves do students enact during this sensemaking to achieve resolution?

A. Video Data Collection Considerations

Given our theoretical framework and the IPLS course setting, our video data collection plan was tailored to the project. We aimed to capture the details of the sensemaking process while the students were collecting their data and analyzing their data, including their body positions, facial expressions, and gestures. We wanted to capture their discussions throughout the lab, knowing that the lab is a noisy, and sometimes chaotic, environment. Given these goals, we placed several external cameras on high shelves above the lab benches. The cameras had wide-angle lenses that were able to capture the entire workstation, including students’ heads and body positions around the computers and microscopes. The cameras had external Bluetooth mics that were placed at the center of student workstations near computers, as this was where students congregated to capture their conversation. The backup audio recorders were next to the Bluetooth mics, positioned for optimal collection of small group discussions.

We also wanted to collect video data from multiple student groups and multiple labs to capture some variety in their sensemaking. Knowing this, we wound up collecting data from 13 groups (N=38 students), each group enacted 4 labs, and each lab lasted for two or three weeks, typically 3-5 hours of video per week for each group. It varied due to the timing of other lab activities, such as warm-up activities and time allocated for writing lab reports. We also need a record of their computer screens to track the details of their data analysis in a spreadsheet. Specifically, we wanted to track how they arranged and manipulated numerical information on their spreadsheets, we wanted to see the different graphs created and how they modified them over the lab class. Thus, we collected this data using screen capture software that ran in the background of the lab computers. We also wanted to triangulate across video and written data with written work. Therefore, we collected their final lab reports, which were written in a scientific argumentation format. Finally, we didn’t want to miss any data due to equipment or human error, therefore we were checking and charging batteries regularly and carefully moving the video files off the cameras and lab computers regularly.

B. Results and Limitations Given the Video Data Collection

The results found that students engaged in sensemaking to resolve conceptual and procedural inconsistencies. We identified instances when students recognized conceptual inconsistencies (identified in video by visual cues such as throwing one’s hands up in frustration, facial expressions of discontentment/confusion, etc.) and engaged in sensemaking by juxtaposing their hypotheses and evidence and then constructing new scientific explanations (identified in audio by quick back-and-forth discussion between students, coinciding with self-questioning, interactions with TAs, and elements of mechanistic reasoning). Comparably, we identified instances when students a procedural inconsistency involved and engaged in sensemaking by proposing and testing a series of causes toward modifying experimental procedures or apparatus. We argued that both types of inconsistencies are generally productive, evident through video by students’ demeanor becoming more positive and affirming, as well as through audio data which picked up their consensus building and descriptions of troubleshooting success and new experimental ideas. Our ability to capture these findings was due to having the video of their entire sensemaking process during the lab, along with the screen capture video of exactly how they analyzed their data in a spreadsheet. Our video was detailed enough that we could capture nuances in their facial expressions, body positions, and gestures that allowed additional evidence for their inconsistencies. If we had only collected written observational notes or only audio data, or more limited video data, these results would not have been discernable.

However, there were challenges and limitations of the data we collected and that had ramifications. We could not analyze certain instances in the data when students interacted with peers in another group or moved out of range of the audio recorders. Finally, we struggled to connect the two streams of video data, from the external cameras and screen capture. There were times we saw students on the external videos referring to something important on the computer screen but couldn’t sufficiently discern what it was via the screen capture software. We missed some data when students took
notes or enacted data analysis on their personal computers rather than the classroom computers with the screen capture software, despite the instructor’s encouragement to use the classroom computers. We also had a few instances where we missed audio data because the audio quality wasn’t sufficient to parse among many simultaneous voices. These limitations impacted our results, but in mitigating them, an important factor was the researcher also being the lab TA. In this capacity, he was familiar with the students’ voices, which in turn supported his subsequent video analysis across these two video data streams and the external audio. As the TA, he had strong expertise with the Physics content and was familiar with the lab nuances, all of which facilitated a detailed analysis of the video.

IV. DISCUSSION: VIDEO DATA COLLECTION DECISIONS ARE IMPACTED BY THE RESEARCH GOALS

Across these two cases, we showed that key decisions made before and during data collection can have profound implications on the data collected and the larger research project. Across the two cases, there were differences in the learning environments (whole class video data gathering versus small group video data gathering) and differences in the length of time (one class period versus a semester). The second case also included significant lab equipment and the use of screen capture software, while the first case included important embodied actions. Importantly, in both cases, details of the collected video were tailored to the specific research goals, assumptions, and theoretical background, and limitations of the data had ramifications on the results. In this article, we took a microscope to the common practices of gathering video data in PER to provide a window into how and why PER scholars made those key decisions. Our goal has been for the PER community to gain a stronger understanding of the important research decisions that are so often kept behind the scenes, and thereby begin to open the black box of research planning and decision making to encourage more transparency across the community.

Moving forward, for PER scholars, we suggest careful consideration of their goals, research questions, and theory and how those factors can impact camera placement and amount of video to collect.

- What is the phenomenon you are interested in? For instance, based on your goal, research questions, and theory, one could potentially be interested in, for instance, whole class sensemaking about certain physics content, individual student interactions with the lab equipment, or instructor’s open-ended questioning during small group work. Identifying a clear phenomenon of interested is important for subsequent data collection decisions.
- In what instructional activity or learning environment does the phenomenon appear? For instance, perhaps the phenomena of interest occurs in small group discussions, whole class activities, lab courses, interviews, or lecture environments. It’s possible the phenomena of interest is more likely to occur in certain courses with certain goals or instructional approaches, or perhaps it occurs in a wide-range of courses. Careful consideration of the learning environment is important for justifying ones choices and not inadvertently collecting data in a setting of convenience.
- How long do you need to record to capture the phenomenon? Depending on the phenomenon and specifics of the instructional activity, one might decide to collect data for a single session, multiple sessions, or even a whole semester. Importantly, one might need many instances of a phenomena, or one might need only one couple instances, depending on the specifics of the research goals. For example, if one is focused on how conceptual knowledge about a specific topic changes over time, then likely many course sessions are an appropriate length of time.
- How will you set up cameras/recording equipment to capture that phenomenon? Based on answers to the prior questions, it may be important to use multiple cameras or possibly a single camera will suffice. For instance, if one is interested in interactions among many individuals in a large lecture environment then multiple cameras would be needed to capture all the individuals, possibly arranged at different sides of the room. Comparably, if one is interested in gestures or facial expressions of a small number of people, then the camera needs to set up in way to capture those details. As another example, if one is interested in a small number of people within a larger learning environment (e.g. small group work within a lab section or lecture), then the camera position needs to account for those individuals within the class. In many of these learning environments, there are likely to be overlapping conversations in a noisy room, and this reality needs to be accounted for in the recording equipment setup.

For PER scholars embarking on new research projects that will rely on video data collection, we suggest careful consideration of these questions in order to implement careful data collection plans. Answers to these questions are going to be impacted by one’s goals, research questions, and theory, but systematically articulating ones answers to these questions is likely to support the optimal amount and type of data for a given project.

ACKNOWLEDGMENTS

The authors thank three anonymous reviewers for their helpful comments. This material is based upon work supported by National Science Foundation Grant No. 938721 and NSF-GRFP via Grant No. 1747505. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.
Shifts in students’ responses to conceptual questions after a new physics conceptual worksheet: Preliminary findings

Lauren Bauman (she/her)
Department of Physics, University of Washington, Box 351560, Seattle, WA, USA, 98195-15603

Lisa M. Goodhew (she/her)
Department of Physics, Seattle Pacific University, 3307 Third Ave W, Seattle, WA, USA, 98119-1997

Anne T. Alesandrini (they/them) & Al K. Snow (they/them)
Department of Physics, University of Washington, Box 351560, Seattle, WA, USA, 98195-15603

Amy D. Robertson (she/her)
Department of Physics, Seattle Pacific University, 3307 Third Ave W, Seattle, WA, USA, 98119-1997

Conceptual understanding is one metric that has been historically valued in the assessment of physics-education-research-based instructional materials. Attending to COncceptual Resources in N (ACORN) Physics Tutorials are instructional materials that are based on research identifying common conceptual resources for understanding physics—good ideas or “seeds of science” which can be developed into more sophisticated scientific understandings. For this study, we used pre- and post-tests and classroom video to assess students’ conceptual understanding as they completed an ACORN Physics Tutorial about electric circuits. We present the preliminary results of our analysis in this paper; mainly, students more often answered the post-test questions correctly and relied on the resource current is responsive after the use of the ACORN Physics Circuits Tutorial than before.
I. INTRODUCTION

Physics education research has supported the development and testing of a host of instructional materials, from Tutorials in Introductory Physics, to the Investigative Science Learning Environment, to Physics and Everyday Thinking, to Peer Instruction [1–4]. Not only are these instructional materials based on research about students’ ideas, but they are also often tested in physics classrooms, in order to provide instructors with information about what outcomes might be expected if the materials are adopted. In many cases, one of the instruments used to test instructional materials is a set of conceptual questions about the specific physics concepts targeted by the materials. Often, researchers use pre- and post-assessments that seek to measure shifts in students’ conceptual understanding, and/or video recordings that highlight how students are reasoning as they work through the instructional materials. Findings from this kind of research have been shown to be a factor in faculty adoption of instructional materials [5].

This paper presents findings from a preliminary study of the effectiveness of introductory physics instructional materials called ACORN (Attending to CONceptual Resources in) Physics Tutorials, which focus on attending to and building from students’ conceptual resources—“seeds of science” that can serve as generative input for students’ learning [6–8]. We provide evidence that students more frequently gave answers that relied on the resource current is responsive (and less frequently gave answers that implied that the battery is a constant source of current) after instruction using an ACORN Physics Circuits Tutorial than before. As we describe in Section II, ACORN Physics Tutorials are open-ended, meant to support students in building from their own ideas, and are not structured to scaffold toward particular canonical understandings; thus, the shifts we see in students’ thinking were not explicitly scaffolded by the worksheet, yet were still supported by the structure the worksheet does provide. We describe the instructional intervention, our methods, and our findings, in service of providing preliminary information to instructors who may consider using these materials in their own courses.

II. INSTRUCTIONAL INTERVENTION: ACORN PHYSICS TUTORIAL

This study explores changes in student responses about current flow in electric circuits from pre- to post-instruction using an ACORN Physics Circuits worksheet [9]. ACORN Physics Tutorials are unique in their design: they elicit common conceptual resources about physics topics that have been identified by previous research, and then provide scaffolding to support students in recognizing and building from their own ideas. In the context of circuits, these resources include: current is responsive, voltage drives current flow, resistance limits current flow, and the way the elements are connected in the circuit affects current [10,11]. Because these resources were identified as common in the context of questions like the ones in the ACORN Physics Circuits Tutorial, we expected that at least some students will use these resources to reason about the electric circuits presented in the worksheet. At the same time, we expect that the particular form and frequency of student use of these resources may vary, given the dynamic and context-sensitive nature of resource activation [12,13].

Structurally, the ACORN Circuits Physics Tutorial prompts students to sense-make [14] about a set of electric circuits composed of ideal wires, light bulbs, and a single battery [see Fig. 1]. Many of the questions in this worksheet give students the ranking of the brightness of the bulbs in the circuit and ask them to explain the observed brightness, rather than make predictions. First, students consider a bulb connected to a pair of charged capacitor plates. The bulb briefly lights up, then dims and goes out. Then, the worksheet presents a simple battery/bulb circuit and a circuit with two bulbs in parallel, as shown in figure 1 (a) and (b). The worksheet then presents a simple circuit with two bulbs in series (c), then a more complex 4-bulb circuit (d). For each scenario, students are asked to sense-make about the observed brightness, using the concepts of current and/or potential difference.

![FIG. 1. Examples of the circuits in the ACORN Physics Circuits Tutorial. Students are asked to make sense of the observed brightness in terms of current and then voltage.](image)

Finally, the worksheet asks students to reflect on their answers to the previous questions and articulate a set of rules that they have been (explicitly or implicitly) using to explain the behavior of electric circuits. Students then use their rules to predict the behavior of a more complex 5-bulb circuit (Fig. 2), test their predictions in the PhET DC Circuits simulation [15], and revise accordingly.

The learning goals of this worksheet are that students will be able to: (i) predict and explain the relative brightness of lightbulbs in series and parallel networks, (ii) predict and
explain how changing the number and arrangement of bulbs affects the current through the battery, (iii) predict and explain the current in various branches of a circuit with light bulbs and batteries, and (iv) predict and explain the potential difference across various circuit elements.

FIG. 2. Example of a “challenge question” from ACORN Physics Circuits Tutorial.

III. METHODS

The ACORN Physics Circuits Tutorial was tested in the Fall of 2022 at a large R1 institution in the Pacific Northwest. The tutorial was given in an introductory, calculus-based physics course with three components: lecture, laboratory (hands-on experimenting), and recitation. In the recitation component of the course, students typically work through Tutorials in Introductory Physics [1] in small groups of 3-5 students with the support of graduate and undergraduate Teaching Assistants. The ACORN Physics Circuits Tutorial replaced one of the weekly Tutorials near the end of the term. Prior to the ACORN Physics Circuits Tutorial, students had received lectures on circuits from their instructor and had completed labs on circuits composed of light bulbs, batteries, and resistors.

We assessed the impact of the ACORN Physics Circuits Tutorials in multiple ways: written pre- and post-test questions, video recordings of students using the worksheet in recitation, and surveys that asked students about their perceptions of their learning. The pre- and post-test questions (Figs. 3 and 4) that are the focus of this paper were used to measure conceptual understanding of current. In these questions, students were asked to consider what happens to the brightness of the bulbs and the current from the battery when a bulb is added in parallel to the circuit.

FIG. 3 Pre-test question.

The pre-test was a multiple-choice-multiple-response question that included a free-response explanation of reasoning, to match the style of pre-class assignments that are typical in the course. Multiple-choice response questions were common student answers from pilot tests of the worksheet. The correct answer selections are “leave the brightness of bulb 1 unchanged” and “increase the current that flows out of the battery.”

The post-test question was chosen because it is a similar but more complex scenario than the one used in the pre-test, which lowers the chance that observed gains are attributable to retesting. This question was formatted as a set of free-response questions to match the style of typical post-class assessments used in the course (notably, a style that is different than typical pre-assessments in the course). The post-test question was given as homework in early pilot tests of the ACORN Physics Circuits Tutorial, and students answered in ways that were consistent with our intent to assess conceptual understanding of current and voltage. A correct answer to this question would explain that when the additional bulb is connected, the brightness of the original bulbs are unchanged because the potential difference across and the current through each is unchanged. Adding a new parallel branch must draw more current from the battery, because the current through the first branch is unchanged, and current flows in the added branch; the added branch decreases the overall resistance of the network of bulbs.

FIG. 4. Post-test question.

To focus on the impact of the ACORN Physics Circuits Tutorial on student thinking about current, we gave the pre-test immediately before the Tutorial, but after other relevant instruction (e.g., lecture and lab). The pre-test was administered online via a learning management system and graded for completion, not correctness. The post-test was given as homework directly after the Tutorial, but students had a few days to complete the assignment. Homework was marked for completion and correctness. This study is based on our analysis of 271 matched student responses (58% response rate). The response rate is low because we only considered students who consented to participate in the study, completed both the pre- and post-test, and answered part (a) of the pre-test by choosing an answer choice that
spoke to current. 26 students were excluded from the study because of this latter criterion.

Based on data reported by the university, the demographics of the students enrolled in the course from which our sample was drawn were 8% Latinx and Hispanic, 1% American Indian, Alaska Native, and Native Hawaiian, 26% Asian or Asian American, 3% Black or African American, 40% white, 4% not indicated, and 17% two or more races/other. The university demographics include: 10% Latinx and Hispanic, 1% American Indian, Alaska Native, and Native Hawaiian, 27% Asian or Asian American, 4% Black or African American, 34% white, 3% not indicated, 22% two or more races/other. As a field, we do not yet know what constitutes a representative introductory physics course, which makes it difficult to say whether or not our sample is representative [16].

We coded responses to the pre- and post-test questions to capture what students said would happen to the current when the bulb was added in parallel. We looked for evidence of student use of the resource current is responsive—the idea that current depends on the circuit elements and their arrangement—in students’ pre- and post-test responses. For the pre-test question (Fig. 3), the coding scheme (Table I) was constructed based on students’ selections for the boxes in part (a) applicable to current. For example, if a student chose “Increase the current that flows out of the battery,” “Decrease the current that flows out of the battery,” or “It will change the current out of the battery somehow, but we need to know the voltage of the battery and the resistance of the bulbs to answer,” we coded it as “Current is responsive.” If they chose “Leave the current out of the battery unchanged,” we coded it as “current stays the same.” This coding scheme was then applied to students’ free responses in part (b) of the post-test question (Fig. 4). Table I summarizes the coding scheme and shows examples of students’ responses from the post-test. In this analysis, we used the lens of resources theory [6-8], which emphasizes the generativity of student thinking, to focus on the potentially productive idea that “current is responsive”; we did not attend to whether students’ responses were completely correct or included accurate reasoning.

IV. RESULTS

The results from our analysis are summarized in Table II. The response “current stays the same,” consistent with the idea that the battery is a constant current source [17–23], was the most popular answer on the pre-test, selected by 56% of students in our sample. It was least popular on the post-test, with only 12% of students in our sample answering this way.

“Current is responsive” includes any indicated change (increase, decrease, current changes but unsure in what way). The frequency of responses that evidenced this resource was 44% on the pre-test and 88% on the post-test. There was a shift from “current stays the same” to “current is responsive” among 47% of the students’ responses in our sample.

Though “current is responsive” includes multiple possible changes to the current, 92% of the post-test responses in this category indicated (correctly) that current increased. This shift from “current stays the same” to “current is responsive” is significant, using the McNemar test for paired, nominal data (X² = 102.382, df = 1, p-value<.0001).

A central tenet of resources theory is that student reasoning is context-dependent, and that even if we observe a student using an idea in one context, it does not mean they will use that same idea in another [12,13]. This complexifies claims about shifts in student responses as evidence of learning; if by learning we mean a stable change in students’ understanding. What we feel we can say here is that student responses more frequently use the resource current is responsive after instruction using the ACORN Physics Circuits Tutorial than before. We hypothesize, then, that the ACORN Physics Circuits Tutorial cues the activation of this resource. This may seem like a meager claim when our goals are often for students to learn and then be able to apply

<table>
<thead>
<tr>
<th>Code</th>
<th>Example response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current stays the same.</td>
<td>“Remain the same. Current out of battery are not influenced by the fifth light bulb.”</td>
</tr>
<tr>
<td>Current is responsive.</td>
<td>“The current would increase because the additional path through the 5th bulb. The resistance would also decrease.”</td>
</tr>
<tr>
<td>Includes: current increases, current decreases, current changes but unsure in what way.</td>
<td>“The current of the battery will decrease because the resistance increases.”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Post-test (N=271)</th>
<th>Pre-test (N=271)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current is responsive.</td>
<td>Current stays the same.</td>
</tr>
<tr>
<td>111 (41%)</td>
<td>9 (3%)</td>
</tr>
<tr>
<td>Current stays the same.</td>
<td>127 (47%)</td>
</tr>
</tbody>
</table>
particular conceptual understandings in new contexts over an extended period of time. Yet, it still feels encouraging to us, in the landscape of instructional materials testing in PER.

V. DISCUSSION & LIMITATIONS

Our analysis shows that the frequency of student responses that drew on the resource current is responsive increased significantly from pre- to post-instruction using the ACORN Physics Circuits Tutorial. However, analysis of written pre- and post-test responses does not allow us to hypothesize about what, if anything, about the instructional context may have supported these shifts. However, we also video-recorded students working through the Tutorial, and we are in the beginning stages of analyzing this data to understand how the worksheet may facilitate learning. Although the primary focus of this paper is the pre- and post-test analysis, we are intrigued by our video observations so far as they lend insight into a possible mechanism for changes in the frequency of student responses that rely on the resource that current is responsive.

For example, we have noticed that students often articulate a “vexation point” [14]—a critical moment when students attend to and articulate an inconsistency or gap in their understanding, the thing that doesn’t “make sense” to them—around the battery being a constant source of current. As predicted by literature on sense-making, these vexation points are often followed by lengthy discussions with tablemates and instructors, usually resulting in students articulating shifts in their thinking.

For example, one student said, “I didn’t know the battery could spit out as much current as it needed!” after completing the sequence described in Section II, Fig. 1. In this quote, we see the resource current is responsive in the way the student explains why the brightness of bulbs A, B, and C in Figure 1 are the same because the battery can “spit out” as “much current as it needed.”

In another example (from a different group), two students participate in this exchange:

Student 1: “Okay, I did not know that the batteries could be the same but have different current.”

Student 2: “Me too… When did we learn this?”

Again, we observe the same “activation” of the resource current is responsive when Student 1 recognizes that identical batteries can “have” different current. Although they are still framing current as a property of the battery, there is a shift in thinking about it as “fixed” versus “variable.” Student 2 shares in the same “a-ha” moment, recognizing this fundamental shift and wondering “when they learn[ed] this?”

We also noticed thoughtful questions and sustained sense-making about current evolve from students’ first articulation of the resource. For example, one student (in a different group), wondered, “Can it [current] know? [...] Can it tell there is gonna be more resistance on this path and know to send it less current through that? Or does it just start going through each one and then realize it has to do more?” In this example, the student poses questions to try and understand how current “responds” to changes in the circuit; can it tell, or does it change once it reaches a certain point? Additionally, this student connects the current is responsive resource to the resistance limits current flow resource, recognizing that more current is needed when there is more resistance.

It is promising that some students are engaging in extended sense-making that seems to cue the current is responsive resource for many students both as they complete the worksheet and after (as evidenced by the post-test analysis). We find how this process is playing out in the classroom somewhat surprising because this worksheet was not designed to elicit any particular difficulties, including the battery as a constant source of current difficulty; rather, the worksheet was designed to support students in explaining a set of observations and then reflecting on the ideas they are already using to articulate and test a model. We have open questions about whether the worksheet is functioning like an Elicit, Confront, Resolve (ECR) sequence [24], a common instructional strategy used in other instructional materials (e.g., Tutorials in Introductory Physics), or if it is functioning differently. We plan to pursue this question in more depth.

Finally, a limitation of our work is that we do not know how representative our sample is, since as a field we do not know what constitutes a representative sample of introductory physics students [16]. However, Kanim and Cid have shown that PER has oversampled from white, wealthy, high-mathematics-SAT-scoring populations of students and then treated this group (implicitly or explicitly) as representative of all introductory physics students [25]. This practice sets up an implicit norm against which all students are measured, even as many such students’ needs and strengths are not considered in the development of research-based materials and insights. It is important, then, to contextualize our findings as coming from the particular population of students we name in Section II.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under grant numbers 1914603 & 1914572. The authors wish to thank Paula R.L. Heron for the helpful conversations and feedback that have shaped our thinking.
REFERENCES


Analysis of student essays in an introductory physics course using natural language processing

Amir Bralin
Department of Physics and Astronomy, Purdue University,
525 Northwestern Ave, West Lafayette, IN, 47907, U.S.A.

Jason W. Morphew
School of Engineering Education, Purdue University, 610 Purdue Mall, West Lafayette, IN, 47907, U.S.A.

Carina M. Rebello
Department of Physics, Toronto Metropolitan University, 350 Victoria Street, Toronto, ON M5B 2K3, Canada

N. Sanjay Rebello
Department of Physics and Astronomy / Department of Curriculum & Instruction, Purdue University, West Lafayette, IN, 47907, U.S.A.

We analyzed the essays that were written on various topics in an introductory physics course using two unsupervised machine learning algorithms. One of them was Latent Dirichlet Allocation (LDA). This algorithm is used for extracting abstract topics from a collection of text documents. The other algorithm was Non-negative Matrix Factorization (NMF). It is used for similar purposes but also in other domains such as image recognition. We applied these two algorithms to the dataset that consisted of $N = 683$ student essays. Although there were some built-in, important differences between LDA and NMF, they both found similar topics in our data by large. This offers instructors a promising and productive way of accessing useful information about their students’ written work, especially in large-enrollment classes.
I. INTRODUCTION

This work contributes to the development of the approach to analyze student work using Natural Language Processing (NLP). The idea is to make machines read and process any text produced in a physics classroom and let humans engage in deeper analysis of that text. For example, students may write their solutions to a given physics problem. If the number of students in the class is too high, then the teacher will likely not have an adequate amount of time to evaluate them all in detail. To assist the teacher, one of the tasks which the NLP algorithms can do is text classification: each piece of text can be assigned a numerical value (e.g. 1 for being correct and 0 for being incorrect) or a categorical label (e.g. “kinematics”, “statics”, and “dynamics”). Other NLP-performed tasks include summarization, sentiment analysis, language translation, and more.

In this study, we focus on Topic Modeling – the NLP task to find themes or patterns within a body of text. Algorithms that can perform this task were developed during the 1990-2010s. A concise review can be found in Blei (2012) [1]. One such algorithm is called Latent Dirichlet Allocation (LDA). It was published by Blei et.al. (2003) [2] and has become popular since. LDA models human text statistically. Consider a text file or document as the collection of words. Each word is not random, but is associated with the topic of the document. The problem is that we don’t know this topic a priori. Moreover, we have a collection of many documents written on various topics. That is why they are referred to as “latent” or “hidden”. Instead, we know the posterior fact – the words produced by each topic. Using probability theory, this problem can be solved with Bayes’ theorem. The name Dirichlet (1805-1859) is present in the name of this algorithm because it uses the Dirichlet distribution for inference and parameter estimation. LDA works best for large corpora of text documents. For example, 5577 articles published in the journal Science Education were analyzed with LDA in Odden et.al. (2021) [3]. A similar work but with a smaller dataset – 1302 papers from PERC Proceedings – was done earlier, by Odden et.al. (2020) [4]. A related work which had a comparable data size with ours is Geiger et.al. (2022) [5]. Its authors analyzed student responses to a conceptual question about the electric circuits in an introductory physics class. Using LDA, they obtained some key student ideas that were common to many of those responses. This opened a way to quickly generate useful insight into student thinking. In addition, there was a recent publication by Wilson et.al. (2022) [6]. The authors analyzed student responses to a conceptual survey using NLP techniques other than LDA. They achieved good agreement between their NLP-generated results for categorizing student responses and the ones obtained by traditional means.

To expand upon this line of research, we implemented LDA in our context and did the same for another NLP algorithm called Non-negative Matrix Factorization (NMF). It was developed in the 1980s under different names, but gained popularity and its present form in Lee & Seung (1999) [7]. Like LDA, NMF also models human text simply as the collection of individual words (also known as the “Bag-of-Words” approach in NLP). However, it arranges these (digital) words in a structured, tabular format – a matrix – and then uses the standard techniques of linear algebra. This makes NMF qualitatively different from LDA as the results (topics) are determined by the data (words) alone, without any statistical distributions involved. Our research questions are: (1) what common topics can be found by both LDA and NMF in our data? (2) what are the differences between the results of these two algorithms?

A. Course context

This study took place at a land grant U.S. Midwestern university. All data came from over 2,000 undergraduate students enrolled in an introductory calculus-based physics course for engineers during the Fall 2022 and Spring 2023 semesters. The student population of the course included 25% women, 10% underrepresented minorities, and 8% international students. In terms of specialization, approximately 80% of the students were majoring in engineering.

The course was titled “Modern Mechanics”. Based on the textbook “Matter and Interactions” (Vol.1) by Chabay and Sherwood [8], its content was centered around three fundamental concepts of classical mechanics: Momentum, Energy, and Angular Momentum. The course consisted of 3 components: lecture, recitation, and laboratory. Lectures were conducted in a traditional format: during two 50-minute-long sessions per week, in a large auditorium. Recitations served as the problem-solving sessions where the students worked in groups to solve problems similar to their exam questions. The laboratory setting was the source of data for our research. All students were divided into groups of three and each group worked as a unit. At the end of each 110-minute-long lab, one person from each group submitted the lab work on behalf of the group. No data thus collected included any personal, identifiable information.

B. Research context

During the last five weeks of the course, students were given a lab assignment: to find a real-world problem related to the course content and write an essay about how to solve it. The writing instructions were provided gradually over the span of the assignment: students started brainstorming about the problem during the first week, then came up with a solution during the next week, and iterated through this problem-solving cycle until the final week of the course.

Computation was an integral part of the labs. Our students were taught the basics of programming (such as variables and loops) using Python. During every lab session, the students also engaged in simple data analysis using the standard computational libraries NumPy, pandas, and matplotlib.
As for the lab notebooks where they had to report their results, the Jupyter Notebook format was used. It is a web-based interactive writing platform which incorporates computer code alongside text. This allowed our students to create professional-looking lab notebooks that integrated both their written work and calculations seamlessly. A snippet from one student essay is shown on Figure 1 as an example.

**FIG. 1.** A fragment from an example essay written by the students for their assignment. Note that there are short code cells alongside normal text with some equations.

II. METHODS

There were two algorithms that we applied to our data and compared: Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF). Both are widely used for Topic Modeling: searching for common sets of words within a text by a computer. The usage of the term set above is important. In computer science, the key property of a set of items is that the order of those items doesn’t matter. In contrast, a collection of items with some order defined for those items is called a sequence. The algorithms and data structures of Topic Modeling operate with sets, not with sequences. This means that when analyzing a human-generated text, both LDA and NMF ignore the order of words and sentences (the Bag-of-Words approach) in that text. This is an important assumption on the part of these models.

A. Topic Modeling

The workings of LDA were well-described in the aforementioned works of Odden and colleagues (Ref. [3], [4]). This algorithm represents the final product of a decade-long effort in the field of latent semantic analysis. The mathematical foundations of LDA are beyond the scope of this work. We only emphasize that it is a statistical model. As such, it needs a lot of data. Searching for an alternative that can remedy the shortage of data, we found NMF. In the original work [7], the authors Lee and Seung showed that it was able to extract certain semantic features from encyclopedia articles. The example they gave was the set of words {president, congress, power, united, constitution, amendment, government, law} which represents a clear, coherent topic.

**FIG. 2.** An example of how vectorization works. Three sentences are given as input. After breaking each sentence into individual words (excluding articles and prepositions), it can be rewritten as a table column or a vector. Combining all three vectors corresponding to the three separate sentences results in the matrix corresponding to the original text. Here, all empty matrix elements are assigned number 0 while the original words remain the same for convenience. In fact, each unique word from the text is transformed into an integer number in the final matrix.

NMF belongs to the family of algorithms used for factor analysis. An algorithm from the same family that was widely used in PER and Science Education is Principal Components Analysis. In the context of NLP, factor analysis becomes relevant when the text data that we want to analyze takes the tabular, matrix form. This can be done by constructing the matrix $V = \{v_1, v_2, ..., v_n\}$, where $v_i$ is the text of the $i$-th document ($i = 1, ..., n$) in the vector form. This is called “vectorization” or tokenization and is shown on Figure 2 as an example. Here, $n$ is the total number of documents in the dataset. Then, NMF finds the simplest possible model of $V$: the product of two other matrices $W$ and $H$, by performing matrix decomposition and minimizing the difference (Euclidean distance or Frobenius norm) between $V$ and $WH$. The matrix $V$ is called “visible” because it contains the actual, observable data. The matrix $H$ is called “hidden” because it contains the values which we cannot observe directly. The matrix $W$ contains the “weights” of the visible variables with respect to the hidden ones.

B. Data Processing

In total, $N = 683$ student essays from two semesters were collected. All essays were retrieved from the university’s online learning platform. No limitation on the size of the essay was explicitly stated. The distribution of the essay lengths measured by the total number of words is shown in Figure 3. On average, an essay was approximately 1,000 words long.

Our data cleaning included the following: (i) discarded all lines of code written by the students in order to only work with text; (ii) discarded all numbers and mathematical symbols in order to keep only the text in English; (iii) parsed the remaining raw text into individual words;
FIG. 3. The histogram of essay lengths measured by the word count. The total number of essays was \( N = 683 \). The number of bins used for plotting was 30. The mean and standard deviation were \( \mu = 1,170 \) and \( \sigma = 595 \), respectively.

(iv) disposed of articles, prepositions, and pronouns among those words in order to keep nouns, adjectives, and verbs. Our data analysis was done in Python. Specifically, we used the library scikit-learn [9] for Topic Modeling. Both LDA and NMF algorithms were accessed through its `sklearn.decomposition` module. The intermediate step between data cleaning and analysis was vectorization (see Section II A): converting the essay words into digits that a model (LDA or NMF) can operate with. This was done using scikit-learn’s CountVectorizer (for LDA) and TfidfVectorizer (for NMF) sub-modules. The Python code that served as the basis for our analysis was taken from scikit-learn and can be found at Topic Extraction with NMF and LDA.

III. RESULTS

First, we modeled our data with LDA. We assigned a fixed number of topics to the LDA algorithm and it found them from the essay dataset. To evaluate its precision, we plotted the associated error vs the number of topics on Figure 4. As depicted, LDA reached a limit at \( n = 20-30 \) after which, increasing this number did not result in any significant improvement. Then, we modeled our data with NMF. The corresponding plot of the associated error is shown on Figure 5. It didn’t reach any limit by increasing the number of topics assigned to the model. Rather, it decreased monotonically. This is a built-in property of NMF. In principle, its error would reach 0 if the number of topics was equal to the number of essays in the dataset. Such perfect precision, however, wouldn’t help with accuracy.

The \( n = 10 \) topics found by LDA are given in Table I. They are represented by the five most likely words associated with each topic. For example, Topic 9 reads \{“wheel”, “friction”, “mass”, “braking”, and “power” \}. From this set of words we can understand or infer a that this topic is about cars. Possibly, about ensuring the safety of a car’s braking mechanism by studying the friction between its wheels and the ground. In principle, there are as many words associated with each topic as there are words in the entire dataset. However, a meaningful LDA topic may be represented by a few words which have the highest probability of occurring in an essay. By glancing at other columns in the table, we can see that not all of the topics are as clear. For example, under Topic 7, there is a word “shirt” that doesn’t fit the context of that topic. In addition, there are several words (such as “ball”, “earth”, and “car”) that appear in multiple columns. This further complicates the interpretability of LDA.

The \( n = 10 \) topics found by NMF are given in Table II. Although some words (e.g. “momentum”) are still found in more than one column, they are all general terms expected to be present in multiple topics. More specific terms such as “ball”, “earth”, and “car”, which have clearer associations in the human mind, are now localized: each one appears in a single column.
TABLE I. \( n = 10 \) topics modelled by the Latent Dirichlet Allocation (LDA) when applied to the essay data. Each column represents a topic, defined by the 5 “top words” that are most associated with it.

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
<th>Topic 4</th>
<th>Topic 5</th>
<th>Topic 6</th>
<th>Topic 7</th>
<th>Topic 8</th>
<th>Topic 9</th>
<th>Topic 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>car</td>
<td>bungee</td>
<td>spring</td>
<td>car</td>
<td>ball</td>
<td>drone</td>
<td>pool</td>
<td>rocket</td>
<td>wheel</td>
<td>rollercoaster</td>
</tr>
<tr>
<td>speed</td>
<td>mass</td>
<td>constant</td>
<td>collision</td>
<td>angle</td>
<td>time</td>
<td>cue</td>
<td>earth</td>
<td>friction</td>
<td>loop</td>
</tr>
<tr>
<td>distance</td>
<td>change</td>
<td>change</td>
<td>distance</td>
<td>mass</td>
<td>collision</td>
<td>mass</td>
<td>mass</td>
<td>cart</td>
<td></td>
</tr>
<tr>
<td>change</td>
<td>speed</td>
<td>mass</td>
<td>ball</td>
<td>height</td>
<td>drag</td>
<td>ball</td>
<td>change</td>
<td>braking</td>
<td>height</td>
</tr>
<tr>
<td>wall</td>
<td>time</td>
<td>height</td>
<td>mass</td>
<td>air</td>
<td>air</td>
<td>shirt</td>
<td>train</td>
<td>power</td>
<td>friction</td>
</tr>
</tbody>
</table>

TABLE II. \( n = 10 \) topics modelled by the Non-negative Matrix Factorization (NMF) when applied to the essay data. Note that the order of NMF topics is not the same as that of LDA nor the topics themselves are identical in two tables.

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
<th>Topic 4</th>
<th>Topic 5</th>
<th>Topic 6</th>
<th>Topic 7</th>
<th>Topic 8</th>
<th>Topic 9</th>
<th>Topic 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>car</td>
<td>rollercoaster</td>
<td>ball</td>
<td>energy</td>
<td>rocket</td>
<td>spring</td>
<td>train</td>
<td>drone</td>
<td>braking</td>
<td>wheel</td>
</tr>
<tr>
<td>collision</td>
<td>loop</td>
<td>angle</td>
<td>force</td>
<td>asteroid</td>
<td>bungee</td>
<td>stop</td>
<td>parachute</td>
<td>force</td>
<td>turbine</td>
</tr>
<tr>
<td>force</td>
<td>cart</td>
<td>player</td>
<td>device</td>
<td>satellite</td>
<td>bullet</td>
<td>energy</td>
<td>terminal</td>
<td>rider</td>
<td>energy</td>
</tr>
<tr>
<td>energy</td>
<td>ride</td>
<td>velocity</td>
<td>angular</td>
<td>fuel</td>
<td>toy</td>
<td>force</td>
<td>delivery</td>
<td>stop</td>
<td>water</td>
</tr>
<tr>
<td>change</td>
<td>height</td>
<td>hit</td>
<td>torque</td>
<td>earth</td>
<td>energy</td>
<td>hill</td>
<td>force</td>
<td>brake</td>
<td>wind</td>
</tr>
</tbody>
</table>

IV. CONCLUSION

This study aimed at contributing to the methodology in PER that is based on using the computational resources offered by Natural Language Processing (NLP). To this end, we performed Topic Modeling on student essays in an undergraduate physics course about introductory mechanics. We used two different algorithms for the same dataset: Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF). The former is a statistical model of digital text treating each topic within that text as a distribution of text’s words. The latter is a model of digital text which treats it as a matrix of text’s words and performs factor analysis on it. Both models offer a quick and powerful way of interpreting class materials: not just essay documents, but any other text files that need to be read and processed. These can be homework assignments, quiz responses, or the answers to some open-ended exam questions. In our context, a teacher may see common themes among the essays prior to reading them. This may save some time and effort in getting familiar with the student work and grading it later.

We found that for our dataset, in particular, NMF offered a complimentary way of interpreting the topics found by LDA. This method of Topic Modeling is the domain of unsupervised machine learning which is hard to evaluate. We expect that one could combine the two algorithms into a single, unifying procedure: fit the entire dataset on LDA, then fit a subset of the data on NMF to cross-validate the LDA results. Overall, the common themes among the topics found by both LDA and NMF are in agreement with each other. The difference in the interpretability of the topics between the two models becomes evident upon a more detailed evaluation of each topic.

V. LIMITATIONS AND FUTURE WORK

A careful reader will notice that the absolute values of the error rate for LDA and NMF in Figures 4 and 5 are not comparable. This means that we cannot compare the performance of these two algorithms directly, by observing the error decreasing with increasing number of topics. Thus, we tried to compare the topics found by LDA and NMF manually, by looking at the corresponding Tables I and II.

In using Topic Modeling for our student essays, we completely disregarded the code and math symbols that were present in the original data. Neither LDA nor NMF are suitable for analyzing these aspects of language in their default modes. Future work in this direction will require either adjusting these two algorithms accordingly or using other NLP methods such as large-language models.

There is a fundamental discrepancy between NLP-based and human interpretation of written text. The topics generated by a model of choice (whether LDA, NMF, or any other) are not related to the topics that a human being may find in the same body of text. The former are mathematically defined constructs that represent our best estimate of what constitutes a “topic”. The latter are certain concepts in our minds which may or may not coincide with these estimates. Therefore, we do not suggest that this methodology can completely replace the traditional human analysis of educational texts.

VI. ACKNOWLEDGEMENTS

This work was supported in part by U.S. National Science Foundation Grant 2021389.
Investigating the assessment landscape of physics graduate programs

Bill Bridges, Caitlin Solis, Joshua Barron, and James T. Laverty
Department of Physics, Kansas State University

Jacquelyn J. Chini
Department of Physics, University of Central Florida

Rachel Henderson
Department of Physics, Michigan State University

As physics graduate programs adapt to an ever-changing world, it is important to update their practices of assessing students. We are interested in the variety of ways that physics graduate students’ comprehension and progress are evaluated. We conducted a landscape study of university handbooks and websites documenting the ways in which students are assessed throughout their program. These practices were compared among departments to determine how diverse department assessments are. We also compared our findings with recommendations from “Graduate STEM Education for the 21st Century” determined by a committee of the National Academies of Science, Engineering, and Medicine. This work will help to better understand what the numerous institutions across the country consider as necessary practices and requirements for graduate students. Understanding this landscape can provide a resource for graduate programs looking to update their practices and a foundation for further investigations into graduate education within physics programs.
I. INTRODUCTION

Recent work on graduate physics programs has focused on admissions practices. These include the reliability of the physics GRE upon admission as a predictor of success and how rubric-based holistic review contributes to a more diverse population of graduate students. [1–5]. There has been less work investigating practices within physics graduate programs beyond admissions. We know very little about the overall landscape of assessment practices required for a graduate program in physics. Ideally, these assessment practices represent what graduate programs of a university value, such as independent scholarly activity, strong research skills, and writing ability.

In addition, as physics graduate programs face an ever-changing world, they may consider updating their practices of assessing students. This could be done in the hope of addressing concerns from broadening participation of graduate students to updating practices that have not been addressed in many years. Universities expressing interest in this will likely search for common practices in other programs to model their updates after; however, we’ve identified that there is a lack of common understanding of the potentially varying practices of physics graduate programs across the country.

For this study, we explore formal assessment practices through publicly available graduate handbooks and websites of departments. We define formal assessment as any way that physics graduate students’ comprehension and progress are evaluated towards meeting the criteria established by the department or university. We acknowledge that analyzing publicly available data is a limitation of this study. If a practice is not specifically mentioned in the website or handbook, that does not mean the practice does not exist in any capacity. Any publicly available, written documents may not fully describe the institution’s assessment practices.

With a goal of better understanding the landscape of the various formal assessment practices for physics graduate programs, we ground our work in how well these practices align with the National Academies “Graduate STEM Education in the 21st Century” report [6]. This report outlines the expectations and guidelines of STEM graduate programs. In this study, we report preliminary results that address the following research questions:

- What are common assessment practices across physics graduate programs?
- How uniform are these practices across programs?
- How well-aligned are graduate student assessment practices with the recommended expectations and guidelines from the National Academies report?

To accomplish this, in Sec. II we describe the National Academies report more concretely in order to contextual our study. In Section III, we describe the methods of developing the emergent categories of assessment practices. We then report on how uniform those categories are among physics departments in Section IV. We also report on how well these emergent categories assess the recommendations of the National Academy’s report in Section V B. Finally, we connect the preliminary results from this study with the numerous avenues of future work this study has to offer in Section VI.

II. GRADUATE STEM EDUCATION IN THE 21ST CENTURY

The National Academies of Science, Medicine, and Engineering ordered an analysis of graduate education in STEM done in 2015 [6]. This report was completed to determine what an ideal graduate STEM education involves for all stakeholders including graduate students, faculty, and programs. It proposes recommendations and core elements of each aspect of graduate education including the Master’s and Ph.D. degrees.

For the purposes of this paper we will focus on the recommendations for the Doctoral degree which include core competencies, career explorations and preparation for Ph.D students, and structure of doctoral research activities. This list of recommendations, listed in Table II, then provides a useful metric to determine how well the assessment practices of different physics departments match with what the National Academies present as important factors necessary to graduate education. For example, the report cites a core element of a quality Ph.D. education based in scientific literacy, communication, and professional skills as the learning of ethics and norms of scientific enterprise. Then when looking through our data, we look for assessment practices aligning with the idea of ethics training and norms.

III. METHODS

A. Context

The American Institute of Physics (AIP) produced a report in 2020 stating that there are a total of 260 universities and organizations that grant graduate degrees in the field of Physics in the United States [7]. This includes Ph.D., Master’s in Arts or Science, Terminal Master’s, Dual Credit programs, and specialty degrees or certification programs specifically in the areas of pure physics, astrophysics or astronomy, medical physics, or biophysics. As we are concerned with the general landscape of physics graduate programs, this provides a robust list of programs that include those different degrees.

B. Data Collection

For this preliminary study, we began with 60 randomly selected universities from the list provided by AIP including the home universities of the authors. We focused on data that is publicly available. A primary source we commonly used to gather information was the department handbooks of graduate programs when available. In the absence of a handbook or
if the handbook’s information did not appear satisfactory, data was collected from the department websites or handbooks of the graduate college. Based on the definition of formal assessment above, we scaled down the information present for each of the university handbook or website into the distinct categories described below. For example, we collected information on practices such as the dissertation and courses required but not information covering the tuition or insurance of graduate students.

C. Data Analysis

Two researchers extracted data for each university to establish confidence. From the information gathered we determined categories of data including Courses (core and elective), Candidacy Exams (subject exams and research proposals), Dissertation, Training, and Miscellaneous requirements. The authors, spanning three universities, met to discuss category development based on the prevalence of language present in the data. These categories and sub-categories are shown in Table I. The subcategories were then determined by the most common practices that fit within each larger category. For example, if a program lists a course titled Solar System Physics it was placed in the Other Courses sub-category. In the event of multiple universities placing this course as a core course then Solar System Physics itself becomes a subcategory.

IV. RESULTS

Due to implications of different degree programs and semester timelines, we further limited the 60 universities investigated so far to include only physics Ph.D. degree programs in universities operating on semester timelines. Doing so resulted in retaining 37 out of 60 universities. The data for these 37 universities included 19 physics graduate handbooks and 18 university websites. The list of emergent categories are listed in Table I.

The following sections describe the categories, including coding decisions and subcategories when appropriate. In all cases, the number of departments that have explicit language of each subcategory is listed in Table I. The categories include: courses, candidacy exams, dissertation, training modules, and miscellaneous. Courses include the core course requirements and electives when appropriate. Candidacy Exams are split between subject exams and research proposals. The dissertation category covers the writing, presenting, and expectations of the document itself. Training covers any training the department requires students to complete, and the miscellaneous category covers any assessment that does not clearly fit within the other categories.

<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategory</th>
<th>Num.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courses</td>
<td>Classical Mechanics</td>
<td>26</td>
</tr>
<tr>
<td>(Core)</td>
<td>Quantum Mechanics</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Electrodynamics</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Statistical Mechanics</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Mathematical Methods of Physics</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Other Courses</td>
<td>21</td>
</tr>
<tr>
<td>Courses</td>
<td>Number Required (Courses/Credits)</td>
<td>16</td>
</tr>
<tr>
<td>(Electives)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Candidacy</td>
<td>Written Portion</td>
<td>16</td>
</tr>
<tr>
<td>(Subject Exams)</td>
<td>Oral Portion</td>
<td>12</td>
</tr>
<tr>
<td>Candidacy</td>
<td>Written Portion</td>
<td>29</td>
</tr>
<tr>
<td>(Research Proposal)</td>
<td>Timeline</td>
<td>18</td>
</tr>
<tr>
<td>Dissertation</td>
<td>Written Portion</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Oral Defense</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Research Credits Required</td>
<td>9</td>
</tr>
<tr>
<td>Training</td>
<td>Human Subjects</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Learning/Teaching Assistants</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Responsible Conduct of Research/Ethics</td>
<td>6</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>Committees</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Annual Reports</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Journal Articles</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>TA Requirements</td>
<td>8</td>
</tr>
</tbody>
</table>

1. Courses

We divided the course requirements into 2 main subcategories, core courses and elective courses. Each core course was counted once if the program required students to take at least one subject. For example, if a program required a student to take two semesters of quantum mechanics we record it as one count of the department requiring quantum mechanics rather than two counts. The courses that appeared most often in the data formed the subcategories. Single instances of required courses that were common to one university were placed in the Other Courses sub-category.

For elective courses, there were two trends common in the data with programs requiring either a certain amount of courses or credits. For the sake of this study we count each university only once if they include a requirement for students to enroll in any elective course.
2. Candidacy Exams

Subject exams were considered as comprehensive examinations of core subjects offered by the department. These were written and oral tests examining the student’s knowledge on specific or cumulative subjects, such as Electrodynamics and Classical Mechanics. Universities that required any form of written or oral subject exam were counted, including the subjects covered in those exams when applicable.

The more common process of achieving Ph.D. candidacy is through a student’s research proposal. We considered any requirement of a student to present current research and propose a research plan. We counted universities that made any explicit mention of a student presenting research to achieve Ph.D. candidacy, regardless of whether those students also needed to pass a specified number of subject exams.

From the data we see a strong emphasis towards the research proposal as expected. While there are a significant number of departments that require subject exams as part of the candidacy process, the number was not as high as anticipated with many departments offering ways to either bypass or fulfill that requirement in different ways.

3. Dissertation

In this category we counted departments that made any explicit mention of the writing of a dissertation and defending it through a formal presentation. We considered also any mention of students being required to enroll in research credits to finish their dissertation.

We see that many departments mention the need for students to complete a dissertation of original research, yet there is a lack of more information given on what requirements are fulfilled. Something that was not as widespread was for some departments to consider the credit hours devoted to work on the dissertation.

4. Training Modules and Courses

From the data we determined that students may have to complete diverse training opportunities across institutions. We counted each mention of training requirements, including whether the training sessions were single hour modules or full semester long courses.

For example, a department requiring a student to pass a Responsible Conduct of Research (RCR) training module offered by the CITI Program would be coded as a short training module offered outside of the department.

We identified relatively few opportunities for students to be trained in different ways including human subjects, TA, or RCR modules.

The miscellaneous sub-categories identified so far are: committees, annual reports, journal articles, and TA (teaching assistant) requirements.

We place the faculty committees within this category as opposed to Research Proposal or Dissertation as the committees are not unique to either subcategory and likely to maintain its structure between these two milestones.

V. DISCUSSION

A. Preliminary Highlights of Analysis

From the sub-categories of core courses, we see a strong emphasis on quantum mechanics and electrodynamics. We had assumed there would be consistent requirements of the core courses and for many there are. Something notable about this data then is the lack of requirements of a classical mechanics or advanced dynamics course. Many universities in the data had this course as an option, able to be substituted by another such as Statistical Mechanics. A number of universities also required students to take a number of courses not expressed in our categories in Table I. These courses include specialty courses like optics or astrophysics, and could potentially reflect the specialities in research of that department.

We found that the most consistent language was for the writing and presenting of the dissertation. While it was the most consistent, it was also the least defined and structured of the formal practices. Many universities opted to refer to the dissertation as a piece of original research completed by the student, but not provide structure or expectations beyond this.

B. Assessing the Recommendations of the National Academies Report

There are a total of 23 recommendations from the National Academies survey in the section only pertaining to the Doctoral degree. In aligning these recommendations with the practices described in Table I, we identify 3 distinct levels: recommendations that are clearly assessed, those that may be assessed, and those that are not being assessed. Focusing then just on the group of recommendations that can clearly be assessed leaves 9 recommendations from the report. We identify the recommendations that best align with the goals of a practice listed in Table II.

From this comparison, we see a strong emphasis towards the research proposal and dissertation covering many of the elements from the National Academies report. From looking at the university handbooks, there are numerous mentions of a student’s dissertation but not about the type of content it has beyond being a piece of original research. The requirements for the dissertation then are left to be determined by the committee or the norms of that student’s field of study. This
TABLE II. Emergent assessment categories from the data that best aligned with the recommendations of the National Academies survey.

<table>
<thead>
<tr>
<th>Learning Goals</th>
<th>Courses</th>
<th>Candidacy Exams</th>
<th>Dissertation</th>
<th>Training</th>
<th>Miscellaneous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students should develop deep specialized expertise</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students should articulate an original research question</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students should design a research strategy</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Students should evaluate outcomes of research</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students should adopt rigorous standards of investigation</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Students should learn and apply professional norms, ethical</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>responsibilities of scientists, and ethical standards</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students should develop the ability to work in interdisciplinary team settings</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Students should acquire the capacity to communicate</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Students should develop professional competencies</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

The ambiguity of the content and requirements of the dissertation makes it difficult to precisely align the recommendations of the National Academies survey.

Another notable result from this comparison is the mention of students learning ethical standards and norms. Of the 37 Ph.D. programs we found only six mentions of ethics training. However, since our analysis is limited to publicly available information, it is possible more universities require ethics training and communicate those requirements to students through other means.

C. Study Limitations

The results of this study are based on public data. We prioritized graduate handbooks for each university but necessarily used the websites if the handbooks were not available or proved insufficient for data collection. By using this type of data, we continually face the possibility that the data is incomplete or inconsistent with what is actually practiced at each university. As evidenced by nearly half of the Ph.D. programs not having available graduate handbooks, we do not have a consistent form of data to use in this study. In future research, we will corroborate these findings with departments.

VI. CONCLUSIONS AND FUTURE WORK

This analysis represents preliminary results from a subset of data that covers only Ph.D. degrees from universities that operate on semester timelines. We determined a list of assessment practices from this data organized into five main categories: Courses, Candidacy Exams, Dissertation, Training, and Miscellaneous. In comparing these categories among departments, we see a tendency towards many universities having explicit language about the expectations of dissertations, research proposals, and course requirements.

In investigating how well the recommendations of the National Academies report are being assessed in these departments, we found some that were clearly being well assessed and those that were not. Those being assessed are largely focused on the research proposal of the Candidacy Exams and Dissertation categories. Recommendations from the survey report that were not clearly being assessed were more focused on the Courses or Training categories.

In particular, the dissertation holds a lot of ambiguity of specific recommendations and how they are directly assessing the recommendations of the National Academies survey. Investigating this ambiguity and how it connects to those recommendations lies within future work of this study.

As of now we have finished coding only 60 of the 260 total universities in our data. In finishing this coding process, we can begin the numerous comparisons between the university practices and how well all of the recommendations of the National Academies survey align with them. Future work then includes: interviewing faculty to confirm practices at varying institutions, comparing practices between all of the degree types, and the various ways that graduate programs value their students’ time.

ACKNOWLEDGMENTS

We would like to thank the College of Arts and Sciences of Kansas State University for their support. Also thanks to KSUPER for their valuable insights. This material is based upon work supported by the National Science Foundation under Grant No. 2013339.
Investigating context dependence of introductory and advanced student responses to introductory thermodynamics conceptual problems

Mary Jane Brundage  
*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA, 15260*

David E. Meltzer  
*College of Integrative Sciences and Arts, Arizona State University, Mesa, AZ 85212*

Chandralekha Singh  
*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA, 15260*

We use a validated conceptual multiple-choice survey instrument focusing on thermodynamic processes and the first and second laws of thermodynamics as covered in introductory physics to investigate the context dependence of introductory and advanced student responses to introductory thermodynamics problems after instruction. The survey has conceptual problems that incorporate many contexts with the same underlying principles and concepts involving internal energy, work, heat transfer, and entropy. Here we focus exclusively on entropy. This study used data from over 1000 college students in introductory-level algebra- and calculus-based physics courses as well as upper-level thermodynamics courses. In addition to prior research, think-aloud interviews with a small subset of students in which they were asked to answer the survey problems while thinking-aloud were useful for understanding the context dependence of student responses in some situations, and why students may have greater difficulties in some contexts than in others. Here we present analysis of data in multiple contexts reflecting students’ ideas about the change in entropy of a gas in spontaneous/irreversible processes and in cyclic processes. We find that a persistent belief in the constancy of entropy even for spontaneous/irreversible processes is a common difficulty among introductory students across problems with different contexts, while upper-level students had great difficulty across contexts in which identifying entropy as a state variable is important. For example, overall, upper-level students struggled somewhat more than introductory students with the fact that the entropy of the system does not increase, e.g., in cyclic processes after one complete cycle. Our findings using a validated survey confirm the findings of prior research in multiple contexts.
I. INTRODUCTION AND GOAL

Physics is a discipline that is devoted to explaining diverse physical phenomena using just a few basic physics principles. To develop expertise and learn physics effectively, it is essential to unpack the underlying meaning of the abstract principles and concepts to recognize their applicability in diverse situations [1-4]. Research shows that identifying and applying relevant physics principles and concepts involved in different contexts is an important hallmark of expertise in physics. Many physics courses focus on helping students learn to discern the deep similarities between problems that share the same underlying physics principles but have different “surface” features, so that students can transfer their knowledge across different contexts to solve problems. Transfer of knowledge refers to the application of knowledge and skills learned in a given context to other contexts [5-9].

Two physics problems that appear to be very similar to a physics expert because both involve the same physics principle may not necessarily look similar to novice students who are still developing expertise [10, 11]. For example, a study on the categorization of introductory mechanics problems [10] based upon similarity of solutions indicates that experts usually group problems based upon the underlying physics principles while novices are more likely to be distracted by other features of problems, and may group problems based on the surface features such as the inclined plane, spring, or pulley even if the underlying physics principles to solve them are different.

The different ways experts and novices categorize problems may also reflect the different ways in which their knowledge is organized [10, 11]. Research suggests that experts in physics have a hierarchical knowledge structure (schemas), in which the most fundamental physics principles are placed at the top, followed by layers of subsidiary knowledge and details [12-15]. This well-organized knowledge structure facilitates their problem-solving process [12-14]. It also guides the experts to recognize the deep features of the problems and makes the transfer of their knowledge to different contexts easier.

Since helping students recognize the applicability of the physics principles and concepts they have learned and apply them correctly in different contexts is an important goal of physics education for science and engineering students, many research studies have investigated students’ ability to transfer their knowledge to different contexts [16-19]. Cognitive theory suggests that transfer can be difficult especially if the source (from which the knowledge is to be transferred) and target (a problem at hand to which the knowledge is to be transferred) do not share surface features. The source may be the context in which a particular physics principle or concept was learned, while the target may be quite a different context.

The failure to appropriately transfer can often be attributed to the fact that knowledge is encoded in long-term memory (LTM) with the context in which it was learned and the features of the target problem to be solved may not lead to accessing relevant resources in LTM even though the two problems share deep features [1, 20]. Solving problems in new contexts correctly requires unpacking and understanding the applicability of the physics concepts in diverse situations.

Research shows that the robustness of the knowledge structure and the context in which the knowledge is acquired can affect an individual’s ability to apply knowledge flexibly across different contexts [10-14, 16, 17]. For this reason, prior studies have used various scaffolding mechanisms to assist students in learning to transfer their knowledge correctly in different contexts [21, 22]. For example, students can be taught to use isomorphic (similar structure) problems and analogical reasoning to link problems that involve the same underlying principles and concepts [21-25]. Isomorphic problems have the same underlying principles or concepts but have different surface features. To improve transfer of knowledge across contexts, it is important to help students contemplate the applicability of the same physics principles and concepts in different contexts and learn to decontextualize knowledge and store it in their LTM at a more abstract level [1].

When transfer of knowledge in a given context is expert-like and the knowledge which is accessed helps students solve a problem correctly, it can be called positive transfer. However, Lobato posited that transfer should be considered from the perspective of the person solving problems [7]. In her actor-oriented transfer framework, if students transferred knowledge in a manner that was not useful to solve a problem correctly in a given context, this is still a transfer of some knowledge from students’ perspective [7]. We can call this type of transfer a negative transfer [7]. Consistency of positive or negative transfer can be measured by using isomorphic problems and investigating how consistently students perform across different contexts. A consistent positive transfer of their knowledge could signify that students have a good knowledge structure of the underlying concepts and principles, and appropriate knowledge is accessed from the LTM regardless of the contexts and surface features of the problems [1]. If students have the relevant knowledge in their LTM to correctly solve the problems posed, a consistent negative transfer could mean that students are unable to recognize the applicability of relevant concepts and principles and may be getting distracted by some surface features of the problem or may have a strong alternative conception pertaining to some problem features [10, 11]. An inconsistent positive transfer to solve problems could imply that students have some knowledge of the underlying concepts and principles, but that it does not amount to a mastery of the material because the appropriate knowledge is accessed from LTM and applied correctly only in some contexts [1].

Here we discuss an investigation of how introductory and upper-level students access their knowledge about entropy and make use of it in various contexts requiring application of the same underlying concepts. Prior research suggests that both introductory and upper-level students have many difficulties with introductory thermodynamics concepts [26-35]. All upper-level students were once introductory students, so the consistency with which they use their knowledge to solve conceptual introductory thermodynamics problems across various contexts—combined with information about solution
methods typically used by introductory students—can provide insights into the learning process that are potentially helpful to instructors in better meeting the needs of students in their physics courses.

There are many factors that can affect student ability to transfer their knowledge such as the context in which the concept was learned, how similar the surface features of the new problems are to the ones they have solved in the past, and whether certain features of the problems act as distractors. Here are our research questions:

RQ1: To what extent are introductory and upper-level student responses dependent on the problem context for problems with the same underlying concepts related to entropy?

RQ2: How different are the context dependencies of introductory and upper-level student performance on entropy problems across different contexts sharing a common theme?

II. METHODOLOGY

The Survey of Thermodynamic Processes and First and Second Laws-Long (STPFaSL-Long), a validated survey instrument with 78 problems, was used in this research; the instrument focuses on introductory thermodynamics concepts. The details of the development and validation of the STPFaSL-Long can be found in Ref. [36] and the survey can be found here [37]. Most problems on the survey have four possible answer choices; most of the problems dealing with entropy in different contexts have options asking whether it increases, decreases, remains the same in the given situation, or whether there is not enough information. Only 22 out of 78 problems are true/false (T/F) problems. Here we only focus on clusters of problems in multiple contexts in which underlying concepts about entropy are similar.

This investigation on the context dependence of student responses uses survey data after instruction (post-test) in relevant concepts. In particular, the written data analyzed here were taken by administering the survey in proctored in-person classes as a post-test after students had learned the relevant concepts, but before their final exam in the course. Students were given some extra credit for completing the survey. These written student data are from 12 different in-person courses from five different large public institutions and all students completed the survey in class on Scantrons in a 50-minute class period. We discuss analysis of context dependence in the written data from three groups of students: 550 students in the introductory algebra-based (Int-alg) physics course, 492 students in the introductory calculus-based (Int-calc) physics course, and 89 students in their upper-level thermodynamics course. Students in the Int-calc courses were typically engineering majors with some physics, chemistry, and math majors, while students in the Int-alg courses were mainly biological science majors and/or those interested in health-related professions. Students included in the upper-level group were typically physics majors in thermodynamics courses or Ph.D. students in the first-year, first-semester of their graduate program, who had not taken any graduate-level thermodynamics. (Since the survey was administered as a pre-test to this latter group of students, they were presumed to have taken upper-level undergraduate thermodynamics.)

The interview data are from 11 introductory and 6 upper-level students from one institution who volunteered after an opportunity to participate in this study was announced. Each interview lasted between 1-2 hours in one sitting depending upon students’ pace. The interviewed students were given $25 for their participation. The interviews used a semi-structured think-aloud protocol. Students were asked to think-aloud as they answered the questions and were not disturbed except to keep talking if they became quiet. Only at the end did we ask them for clarifications of points they had not made clear.

III. RESULTS

Table I shows responses to problems related to entropy change posed in different contexts. Three distinct student populations are represented: Int-alg, Int-calc, and Upper Level.

Table I. Percentages of introductory algebra-based (Int-Alg), calculus-based (Int-Calc) introductory physics, and upper-level students whose post-instruction responses were in the categories shown for problems probing issues related to entropy in various contexts. The problem number for each problem is shown.

<table>
<thead>
<tr>
<th>Correct answer in bold, difficulties unbolded</th>
<th>Problem #</th>
<th>Int-Alg</th>
<th>Int-Calc</th>
<th>Upper Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entropy of the universe in a spontaneous process increases (Correct)</td>
<td>17</td>
<td>20</td>
<td>23</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>53</td>
<td>21</td>
<td>23</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>50</td>
<td>47</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>43</td>
<td>35</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>67</td>
<td>47</td>
<td>40</td>
<td>83</td>
</tr>
<tr>
<td>Entropy of the universe in a spontaneous process remains constant (Incorrect)</td>
<td>17</td>
<td>71</td>
<td>69</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>53</td>
<td>68</td>
<td>62</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>27</td>
<td>29</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>43</td>
<td>47</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>67</td>
<td>34</td>
<td>38</td>
<td>14</td>
</tr>
<tr>
<td>ΔS=0 after a reversible cycle (Correct)</td>
<td>8</td>
<td>65</td>
<td>54</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>67</td>
<td>63</td>
<td>57</td>
</tr>
<tr>
<td>ΔS=0 after a reversible cycle (Incorrect)</td>
<td>8</td>
<td>19</td>
<td>25</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>22</td>
<td>27</td>
<td>38</td>
</tr>
</tbody>
</table>

a. ΔS=0 for the universe in an irreversible process

Table I focuses on the context dependence of student responses on problems 17, 21, 53, 67 and 75 pertaining to the fact that for the universe, ΔS > 0 for a spontaneous/irreversible process. The fifth problem in this category, problem 67, involves an isochoric process with net heat transfer to an ideal gas; it is explicitly identified as an “irreversible” process in the problem statement, the only one of the five so identified.

Although four of these problems, 17, 21, 53 and 75, have contexts involving isolated systems with spontaneous processes,
they have different surface features that can distract students. For example, Problems 21 (shown in Fig. 1) and 75 can be viewed as similar to each other. Problem 21 involves a free expansion with an ideal gas initially in one chamber expanding into a vacuum when the stopcock is opened, while problem 75 involves two different non-interacting ideal gases initially in separate chambers mixing with each other. The context of problem 75 can be viewed as analogous to the free expansion in problem 21.

FIG. 1. Free expansion of a gas in an insulated container. As shown in problem 21 of the STPfaSL-Long, the gas is initially in thermal equilibrium and confined to the left chamber. When the stopcock is opened, the gas is allowed to expand evenly until the system reaches equilibrium. The problem asks if the entropy of the system increases, decreases, or remains the same after the stopcock is opened and equilibrium is reached.

Problems 17 and 75 are almost identical, but very different from the 21/75 pair which showed many more correct responses. Moreover, while the 21/75 pair had lower error rates than 17/53, the results for 21 and 75 were not as similar to each other as those in the 17/53 pair that shared an identical element. As for why error rates in 17/53 were higher for introductory students than in the other pair, interview evidence suggests that students may view free-expansion as a clearer signal of entropy increase than heat flow.

Table I shows that a majority of the upper-level students provided the correct response regardless of the context of the problem, but there is a large context dependence for both introductory groups, with a large fraction of the introductory students not providing correct responses for each problem context. In particular, upper-level students remain fairly consistent with their responses for all five problems, with percent correct ranging between 78%-87%. This suggests that most upper-level students can consistently correctly transfer their knowledge across different problem contexts involving ΔS>0 for the universe in spontaneous/irreversible processes. Thus, we will now focus mainly on introductory student responses to these problems.

The introductory groups had a common but highly context-dependent alternative conception that ΔS=0 for the universe in the spontaneous/irreversible processes. Table I shows that introductory students struggled the most with problems 53 and 17, as less than 25% of them provided correct responses to either of those problems. (These two problems both involve spontaneous heat transfer from the hot substance to the cold substance; the substances were solids in problem 17 and gases in problem 53.) Furthermore, Table I shows that the contexts of problems 17 and 53 are so challenging that roughly two-thirds of both introductory groups provided incorrect responses stating that entropy of the universe in these processes does not change. Interviews corroborate these findings. For example, on problem 17, one interviewed student who thought ΔS=0 said, “Since there was no loss [of heat] to the environment, we will assume that the entropy has not changed.” On the same problem, another interviewed student said, “Change in entropy must be zero because of equilibrium.”

Table I also shows that for both introductory groups, the correct response rates on problem 21 (free expansion) are only slightly higher than on problem 75 (gas mixing). However, there is a large difference in introductory groups’ incorrect ΔS=0 response. While 47% and 43% of the Int-calc and Int-alg groups, respectively, responded ΔS=0 for the mixing process (problem 75), the corresponding percentages for the free expansion process (problem 21) are 29% and 27%. Response rates for problem 67, the isochoric process explicitly identified as irreversible, fell between those for problems 21 and 75, indicating that even this problem was quite challenging.

In summary, upper-level students performed significantly better overall than introductory groups and were more consistent in their responses. Introductory students’ responses

for the universe. One might then expect similar student outcomes for the 21/75 pair on the one hand, and for the 17/53 pair on the other, but not necessarily similar results when comparing the two pairs. In fact, that is exactly what is found: response rates for the 17/53 pair are almost identical, but very different from the 21/75 pair which showed many more correct responses. Therefore, while the 21/75 pair had lower error rates than 17/53, the results for 21 and 75 were not as similar to each other as those in the 17/53 pair that shared an identical element.

Despite the similarities linking the 21/75 pair and the identical element joining the 17/53 pair, the only feature that the 17/53 pair truly shares with the 21/75 pair is that they are all irreversible. That is, all four problems involve isolated systems undergoing spontaneous processes leading to ΔS>0.
were context dependent, with mixing/free expansion problems being easier for them than were heat transfer problems. The correct response rates of the two introductory groups for all problem contexts related to entropy of the universe in irreversible processes are similar, all in the 20-50% range.

b. In a complete cycle, $\Delta S=0$

Table I also shows the context dependence of student responses on problems 8 and 24 pertaining to the fact that $\Delta S=0$ for a gas that undergoes a full, complete cycle. Problem 8, shows a PV diagram, shown in Fig. 3, with three processes that form one complete counterclockwise cycle while problem 24 shows a PV diagram with four processes that form one complete clockwise cycle, shown in Fig. 4. After a complete cycle, the initial and final states are the same, so there is no change in entropy of the gas. For problem 8, the types of processes in the counterclockwise cycle are not explicitly mentioned, but for problem 24, the specific processes constituting the clockwise cycle are explicitly given.

![FIG. 3. The PV diagram of a gas undergoing a complete counterclockwise cycle in problem 8.](image)

![FIG. 4. The PV diagram of a gas undergoing a complete clockwise cycle in problem 24.](image)

The answer options for these problems were that in one complete cycle, the entropy of a gas would either increase, decrease, remain the same, or there wasn’t enough information. Table I shows that students in all three groups had difficulty with the fact that entropy is a state variable and $\Delta S$ is therefore path independent. In particular, during a thermodynamic process, regardless of how a gas gets from its initial state to its final state, only the two end points determine the change in entropy. In a cyclic process after one complete cycle, since the gas ends up in the same state that it started in, $\Delta S=0$.

 Sheer instruction can assist students in recognizing the applicability of different physics concepts learned in different contexts. However, the possibly slightly worse performance of all three groups on problem 8 compared to problem 24 may be due to the contexts of problems 8 and 24 involving counterclockwise and clockwise cycles, respectively. The correct response rates for problems 8 and 24 for upper-level students were 49% and 57%, for Int-calc group 54% and 63%, and for Int-alg group 65% and 67%. Thus, introductory groups slightly outperformed upper-level students. The comparison of the performances of upper-level and introductory students shown in Table I suggests that there is no significant learning in the upper-level courses pertaining to entropy being a state variable. This is a cause for concern that upper-level thermodynamics instructors should take into account.

The most common alternative conception for all groups was that the entropy of a gas increases after a complete cycle. Table I shows that across both problems, the upper-level students have a stronger alternative conception (approximately one-third of the upper-level students) than introductory students (approximately one-fourth of the introductory groups) that the entropy of a gas increases after one complete cycle. This is echoed in interviews with upper-level students. For example, for problem 24, one upper-level interviewed student said, “is heat in or heat out related to the entropy?” and then tried to determine the heat transferred into the system in each step of the cycle in order to find the change in entropy. It is true that in a clockwise cycle, there is a net heat transfer to the system, but this upper-level student was distracted by the individual processes and reached the incorrect conclusion, not recognizing that entropy is a state variable and is therefore unchanged since the initial and final states are identical after a cycle.

IV. SUMMARY

We find that introductory students’ application of their knowledge about entropy depends heavily on problem context. Introductory students’ responses on irreversible processes were context dependent; they performed better on processes involving gases expanding than on problems involving heat transfer. Interviews suggested that many students are unaware that heat transfer processes involve net entropy increase. For cyclic processes in which $\Delta S=0$, correct responses were given slightly more often by introductory students than by upper-level students; both groups’ correct response rates were in the 49-67% range. Lack of opportunity to develop a robust knowledge structure can prevent even the upper-level students from solving conceptual introductory problems successfully in different contexts. Therefore, one major goal of both introductory and upper-level physics instruction should be to help students construct robust schemas so that their well-organized knowledge structure can assist them in recognizing the applicability of different physics concepts learned in different contexts.
Considering the Departmental Action Leadership Institute as a Community of Transformation: What’s highlighted and what’s missed?

Robert P. Dalka, Chandra Turpen, Diana Sachmpazidi, and Fatima N. Abdurrahman  
*Department of Physics, University of Maryland, College Park, MD 20742*

David A. Craig  
*Department of Physics, Oregon State University, Corvallis, OR 97331*

Joel C. Corbo  
*Center for STEM Learning, University of Colorado Boulder, Boulder, CO 80309*

The Communities of Transformation (CoT) framework is a variation on Communities of Practice that models groups aimed at changing existing institutional practices by challenging underlying value systems. The CoT framework has the potential to provide insight into STEM initiatives designed to promote institutional change. We share results from applying this framework to the Effective Practices for Physics Programs’ (EP3) Departmental Action Leadership Institute (DALI). DALI supports cohorts of physics faculty (change leaders) in leading change efforts in their departments. Change leaders apprentice into effective change strategies though sustained programming while enacting these strategies within their own Departmental Action Team (DAT). Through analysis of interviews with change leaders, we identify ways in which DALI aligns with, and departs from, the CoT framework. We present the results of this initial study to showcase which aspects of STEM change initiatives can be highlighted, and what may not be captured, by a CoT lens.
I. INTRODUCTION

Physics departments are experiencing a wide range of challenges and opportunities that may require them to transform their undergraduate programs [1–4]. Yet, physics faculty do not often have training in leading change efforts. Many may require a support system to meet them where they are and guide them in their efforts [5]. The Departmental Action Leadership Institute (DALI) was designed to bridge this gap between training and needs of physics departments by providing sustained resources and guidance to a community of physics faculty pursuing change efforts in their departments [6]. DALI supports important change work within physics higher education and can be compared to similar models for change. The Communities of Transformation (CoTs) framework models how communities support individuals transforming cultural practices [7]. While DALI is not specifically designed as a CoT, we see an opportunity to explore how the CoT framework can help understand how DALI supports physics faculty in leading change in their departments.

A. Communities of Transformation

It is helpful to consider the Communities of Transformation framework through comparison to similar models. Professional (or Faculty) Learning Communities (PLCs/FLCs) bring a group of practitioners (e.g., college faculty) to meet under the guidance of a facilitator to learn how to improve their practice (e.g., teaching), with a focus on individual growth [8, 9]. PLC/FLC membership often involves a formal sign up process (e.g., an application) and a well-defined time span (e.g., meeting for one semester) [10, 11]. Facilitators play a central role in providing support tailored to the needs of the PLC/FLC members [12]. In physics both in person and virtual PLCs/FLCs have been shown to increase the reach of many research-based instructional approaches [13–15].

Communities of Practice (CoPs) are more organic, where a group of practitioners come together due to shared interest in a topic (e.g., implementing equitable teaching) with the desire to collectively improve their practice [16, 17]. CoP members collectively share best practices and develop new ideas to help each other improve [18]. Members may come and go, but the community can continue to exist indefinitely, oftentimes crossing institutional boundaries. In education research, CoPs have been shown to support the uptake of instructional strategies within STEM [19–21].

A Community of Transformation is a variation on a Community of Practice that is defined by three core elements: (i) challenging existing values/adopter a new philosophy, (ii) creating space for observing and living the new values/philosophy through practice, and (iii) creating a network of peers/community to help enact new practices [22]. These three core elements distinguish CoTs from more traditional CoPs (themselves distinct from PLCs/FLCs). While CoPs and PLCs/FLCs focus on iterative improvement of existing practices, the focus of a CoT is actively transforming and/or creating new cultural practices. Researchers have found that the adoption of a new philosophy by CoT members motivates them to engage in practices in their own spaces that often go against the typical practices of those spaces [22, 23]. Additionally, while CoP networks are organically built and PLC/FLC networks are externally organized, CoT networks feature a center group of members, who may both be intentionally and/or organically developed, who serve as mentors and coordinate communication within the community [24].

Through both their design and the forms of participation of their members, CoTs have been shown to have a high potential for supporting institutional change. CoTs support individual outcomes, such as learning about and improving teaching practice as well as feeling reenergized around this work [25]. They also contribute to transformations within institutions through the work of their members [7]. CoTs provide sustained support, involve multiple individuals from different institutions, build skills in members to communicate their work beyond the community, and identify key leaders who can sustain the community [23]. Given the power of CoTs in creating change, our goal in using the CoT framework to understand DALI is to identify how the three CoT core elements are expressed in DALI’s structure and how the DALI structure may be different from that of a CoT.

B. Departmental Action Leadership Institute

DALI is part of the Effective Practices for Physics Programs (EP3) Initiative, led by the American Physical Society (APS) and the American Association of Physics Teachers (AAPT) [26]. DALI supports cohorts of physics faculty through high-touch mentorship in change practices so they can become better-equipped to lead change efforts in their own departments, such as building stronger retention practices [6]. Articulation of the specific change project is part of the department’s application to the DALI (Participating departments designate two faculty members as change leaders. They attend an initial multi-day DALI workshop and subsequent regular meetings with their cohort for one year. Change leaders are responsible for establishing their own Departmental Action Team (DAT) that they facilitate while receiving guidance and feedback from DALI [27]. The DAT is then responsible for carrying out the change effort within each department. Each DALI cohort is led by two facilitators with experience in departmental change work.

The philosophy behind the DALI model of change is that successful change efforts are deliberately designed, context dependent, driven by a sense of ownership and broad stakeholder engagement, grounded in evidence, and built through an ongoing process [28]. These design choices form a philosophy that change leaders come to adopt over the course of their time in DALI. The DALI provides a community of peers who the change leaders can connect with around shared challenges and ways they work to overcome these challenges.
In our previous research, we have seen how change leaders feel more prepared and confident in their ability to lead change after DALI [6]. In case studies of the DALI-supported DATs, we have seen that the change leaders have facilitated teams that take a collective approach to data collection and sensemaking [29]. Change leaders have come to see students as partners in change work and prioritize their perspectives to better understand the challenges their departments face [30].

In bringing CoT as a lens to understand DALI, we seek to highlight the aspects of DALI that have been shown to be important to supporting change in CoT models. Our two goals are to identify (a) implications for the design of DALI and how to further engage change leaders and (b) aspects of DALI that are different from CoT and why we think those differences are important. We ask the following research questions to guide our analysis:

- What evidence do we see of the central elements of the CoT framework within DALI?
- What important features of DALI extend outside of the CoT framework?

II. METHODS

The data for this study comes from the set of interviews conducted with change leaders in the first cohort of DALI. In this cohort, eight of the ten DALI change leaders agreed to participate in our research efforts. This represents four of the five departments that were a part of DALI. Change leaders represent a wide variety of departmental experience. See Table I for more information on the change leaders.

Over the course of their participation in DALI, our research team conducted three rounds of interviews. For this early analysis using the CoT framework, we have focused our analysis on the first round of interviews, which were conducted about a month after the DALI kick-off workshop. Interviews followed a semi-structured protocol with questions that focused on change leaders’ reflections on their DALI participation, the facilitation of their DATs, and their own thoughts and experiences with change work.

A coding protocol was established from the three core elements of the CoT framework described in Sec. I A. A fourth category was used to identify aspects of DALI highlighted by change leaders that were not captured by the three core elements of CoT. This protocol was used by R. D. to analyze the first round of interviews. In coordination with the rest of the research team, we identified themes that emerged within each category. We use evidence of these themes to show how the CoT framework can highlight certain aspects of DALI and as well as important elements of the DALI model missed in the CoT framework.

III. RESULTS

This section is organized into four subsections, one for each core element of CoT and a fourth for what was not captured by the core elements. Within each core element subsection, we provide examples of how the change leaders’ reflections align with that particular CoT element and note where there are departures from CoT. While there are many different examples that we found within each category, we highlight just one or two in this paper due to space constraints.

A. Adopting a new philosophy

In the CoT model, the philosophy of the community is essential for members to adopt and center in their work to motivate their practice. In interviews, change leaders identify many aspects of the DALI philosophy, including deliberate approaches to change centering data-driven decisions, shared leadership of the DAT that involves multiple stakeholders, building a collective sense of ownership over the change process, and valuing the partnership with student members of the DAT. The DALI philosophy, and the change leaders’ uptake of it, has been written about previously [6, 29, 30], so we will limit the space we dedicate to it here except to note that the DALI approach to change work runs counter to the way that these change leaders have approached change in the past. Here, we will demonstrate how the new philosophy came to be adopted by change leaders by looking at the “change should be deliberately designed” aspect of DALI.

Christina, the department chair of Cypress University, described that she and her change leader partner, Charles, came to DALI to get some “good ideas on how to revitalize our department.” She came in looking for direct answers to what she saw as challenges they faced with the undergraduate program. However, she found that the DALI philosophy was to approach change as an opportunity to better understand those challenges to design more appropriate solutions.

“When I came first to the workshop I was very surprised how slow it was going. But then they told us this is actually an approach we’re going to take […] it was not very coherent with the way- how I wanted to do it. But as I went [...] it kind of sanked in and then I started enjoying it. And seeing like- yeah, you need you don’t need to overload and rush, you just need to take time and understand things.” — Christina
Although this new approach was not what Christina expected, she came to like the philosophy because she found it compelling. Evidence of similar shifts with regard to various aspects of the DALI philosophy was exhibited among all the change leaders within the first cohort. DALI has built a space where change leaders can come to adopt a new philosophy, even when it challenges previously held cultural beliefs.

B. Space to enact new practice

The DALI workshop and the sustained meetings during the academic year serve as a space where the new change practices are modeled by the facilitators. Change leaders apprentice into these practices through their participation. The local DATs serve as a secondary, highly overlapping space where the practices are enacted by change leaders. The movement of practices between these two spaces is central to how DALI attempts to support change work.

Change leaders discuss how they take the practices modeled in the DALI space into their DAT to create a shared sense of ownership within the team. Harold described how he and his change leader partner, Henry, did this for their DAT.

“Henry and I had our first DAT meeting with our Hemlock University group and we used the norm setting exercise that [the facilitators] did. Pretty much mirrored their methodology on that. [...] So I think you know some of these activities that get people involved and interacting from the administration to the faculty to the students, everybody kind of on the same page and doing the same things. I think [it] really helps break the ice and and let us all realize that we’re kind of on an even playing field.” — Harold

Similar to Harold, Misha describes how she enacted the norm-setting activity within her DAT that was modeled in the DALI workshop. She describes a slightly more rocky implementation than Harold did. Misha is “fearful that the power differentials are going to come in again” with regard to a particular faculty member who “said he wants to be in there but he didn’t show up for the meetings.” Due to some of these interpersonal challenges with implementing the new practice, Misha says that “I don’t think we have done a good job, establishing the ground rules. Because we do feel like we might not be established at all.”

In examples from other change leaders, we see how the practices that they use in the DALI meetings are used in the DATs with varying levels of success. It is important for the change leaders to have both of these overlapping spaces, as they value the opportunity to learn in DALI, try things out in their DATs, and come back to DALI for advice on how to move forward. What is not captured well by the CoT model is this intertwining of the DAT and DALI spaces, which can make implementation of these practices tricky for change leaders.

C. Network of peers

The network of peers within DALI is reminiscent of that within the CoT framework. However the way it is structured is more similar to that of an FLC. Membership is determined by the leadership of DALI, and the formal DALI programming is time-bound, lasting about a calendar year. Even with these differences, it is helpful to view the DALI community through a CoT lens as it opens up opportunities to reflect on how the network of peers in DALI supports its members and may inform future development of the DALI network.

Many of the change leaders describe a sense of togetherness that surprised them. Sharing their own experiences of challenges in their departments and their commitment to change is something that brought them together. Paul describes how these connections were built over the workshop.

“I was impressed by what I felt was a sense of camaraderie that developed initially, maybe not the first meeting, by the second or third meeting. Even though it might have developed a lot faster if you were there in person, but I was impressed with the fact that people are comfortable with each other from different institutions.” — Paul

This cohort had their kick-off workshop over Zoom as there were still travel precautions due to the pandemic. Nevertheless, the change leaders were still able to connect in this virtual environment. This type of supportive community makes a space where change leaders are comfortable to share their struggles and find togetherness in addressing similar challenges.

We also see change leaders in these interviews discuss how they benefit from the stories of others in other departments. Henry described what he saw as important about this community and how it tied to supporting their local change work.

“I thought it was important for us to look outside our own walls, and I saw the DALI is an opportunity to tap into what’s happening at other institutions as well as the direction that the facilitators might take us to think about areas that we can continue to grow.” — Henry

Through connections with individuals, Henry hopes to gain insight into other physics departments. Additionally, we see Henry identify the facilitators as being able to provide additional guidance that goes beyond the way he characterizes his fellow change leaders. Similarly, other change leaders position the facilitators as highly influential in supporting the change work. In the CoT model of a network of peers, there is a central group that organizes the community and provides mentorship for newcomers. However, the role we see the facilitators play in DALI is more similar to an FLC model. The CoT framework has highlighted some ways in which DALI builds a community to help enact new practice, but we see that some of the features are different than CoTs.
D. Unique DALI features

The partnerships between change leaders that represent their respective departments supports the translation of DALI practices into the local change work. We have termed this the change leader dyad. Many change leaders reflected about how this was important for them to navigate both the DALI space and in leading their DATs. For Charles, the change leader dyad is also important given the institutional roles that each change leader has.

“I got to meet with my department head, [Christina], to discuss [the workshop]. So it was good to kind of hear what she found interesting from the orientation and I can discuss what I found interesting as well. And it kind of helped us plan our next course of action.” — Charles

There is a sense for change leaders that they are not attempting the change effort alone in their departments. For CoTs, it has been shown that with more members from the same institution, the change work goes further at that institution [23]. However, it is not a central piece of the CoT model, whereas it is for DALI.

In a CoT framework, the change agents would be defined as the community as they are the ones who are a part of DALI. DALI is structured such that change leaders work with DAT members (staff, students, other faculty, etc.) to pursue their change efforts. Through a CoT lens, these DAT members are not captured. On Maple College’s DAT, the student members organized and led a focus group for collecting perspectives on the undergraduate program from other students.

“We’re having the students facilitate [the focus group], because we don’t want the faculty to be there because we don’t want them to worry about like hurting our feelings or the fact that they’re in our classes. So the students are facilitating it and I won’t lie, this is making us both a little bit nervous, [[laughter]] but we’ll see.” — Morgan

The work that students contribute in DATs pushes back against what is typical in departments, leading to Morgan’s nervousness about the process. The inclusion of students on DATs, what they contribute to in the change effort, and the tensions that arise are all important parts of DALI. When just modeling the community of change leaders, we miss this important work by DAT members and how this may be complicated due to institutional power dynamics. The CoT model has been helpful in highlighting particular aspects of DALI, but we find that many of the important, and unique, DALI structures are backgrounded when attempting to frame DALI solely as a CoT. In particular, viewing DALI as a CoT, with its focus on the change leader dyad, overlooks the critical aspect of transformation of departmental culture (not to mention the success of the change efforts led by the DATs in each department) that is central to the DALI model.

IV. DISCUSSION

The DALI philosophy is well established and documented in our other work, where we see how it challenges typical approaches to change work. Change leaders adopt the DALI philosophy themselves even when it went against what they expected. Here, we have used the CoT core elements as a lens to understand some of the features of DALI in more detail.

The CoT framework does not directly model the local institution spaces in which the CoT participants act. However, the DAT is a central focus of the change leader’s work. Each DAT exists within the context of the local institutions and departments, meaning that the ways that change leaders “observe the new practice” are highly influenced by the histories and relationships within their departments. Moreover, there is an important interplay between what the change leaders learn about and share in the DALI and what they enact in their DATs. Thus, it is not possible to completely understand how the DALI functions without looking outside the DALI to understand the DATs.

As with CoTs, DALI helps to create a supportive network of peers for change leaders, building camaraderie across institutional boundaries and energizing change leaders. However, this community differs from typical CoTs in being time-bounded and intentionally-structured by the facilitators. Here, there are opportunities for DALI to borrow from successful CoTs to find ways to keep the cohorts connected after formal DALI activities end as well as to support change leaders who want to move into new mentorship roles. For example, new DALI facilitators have been recruited from earlier cohorts of DALI change leaders.

DALI was not designed as a CoT, yet we have found that using the CoT framework as a lens when analyzing change leader interviews has provided insights into the structure of the DALI. This may help shape the growth and development of the DALI model in useful ways, specifically around how DALI (as well as a growing community of DALI change leaders) can potentially support change efforts over a longer period of time than the initial year of the DALI curriculum. These insights can also help DALI participants better understand how to sustain and institutionalize their change efforts. CoT is a useful framework for researchers and practitioners to use in their work with groups that support STEM institutional change. The unique features of DALI motivate the need for further study in order to develop a model that externalizes the DALI approach to inform physics change initiatives, building from characteristics of CoT and other frameworks.

ACKNOWLEDGMENTS

We thank the change leaders for their participation in this research study and the members of our research group for their insights and feedback. This work is supported by the NSF under Grant No. 1821372. RPD was supported by the NSF GRF under Grant No. DGE 1840340.
Using clusters of models of disabilities to describe support for mentees with disabilities

Constance M. Doty (she/her)
Department of Physics, University of Central Florida, 4111 Libra Dr., Orlando, FL, 32816

Dan Oleynik (he/they)
Department of Physics, University of Central Florida, 4111 Libra Drive, Orlando, FL, 32816

Erin M. Scanlon (she/her)
Department of Physics, University of Connecticut – Avery Point, 1084 Shennecossett Rd., Groton, CT, 06340

Jacquelyn J. Chini (she/her)
Department of Physics, University of Central Florida, 4111 Libra Dr., Orlando, FL, 32816

Students with disabilities involved in postsecondary physics education may benefit from research opportunities and mentorship. However, the literature documenting supports provided by physics mentors to disabled students is limited. In this study, we analyze interviews with five mentors who either instruct physics courses or lead a research group for examples of how they support disabled students doing research or seeking career advice. Furthermore, we contextualize the examples of supports using six models of disability. Models include the cause of disability (medical/social), the effect of impairment on well-being (tragedy/affirmative), and the dichotomy of dis/ability (minority/universal). We find mentors discuss supports provided to disabled students in research settings that align with clusters of models of disability. While there is not one set of models that yields a one-size-fits-all solution, the universal model plus social model cluster can help mentors design useful and durable supports.
I. INTRODUCTION

Students with disabilities are present in all levels of postsecondary education [1], with 10.3% of individuals receiving a research doctorate in physical sciences reporting at least one disability [2]. Mentors can provide support for students with disabilities, such as research opportunities or advice about future careers. Strong mentor-mentee relationships can support the growth of a science identity for undergraduates [3] and positively impact self-efficacy of doctoral candidates [4]. Disabled students have also reported benefits of participating in undergraduate research opportunities [5], and mentorship can help doctoral students with disabilities persist in research activities (i.e., students with depression benefit from positive mentor-mentee interactions [6]). Since mentors play a significant role in who continues to participate in the physics community (i.e., fostering acculturation into the field [7]), it is important to understand their perspective of disability when they support disabled students.

In this study, we use a constant comparison method [8] to analyze interviews with physics faculty from institutions across the United States to expand the literature about mentor support for students with disabilities in physics settings. Disability is a colloquially dynamic word that is context dependent. For example, in some government policies disability might refer to the incapability to work, while in social settings disability might refer to the lack of access to social resources [9]. For this reason, we categorize the examples of mentor supports using a framework that blends models of disability together into a cluster [10]. We aim to answer the following research question: How do the supports mentors provide to disabled students align with models of disability? The goal of this paper is to investigate current support for disabled students provided by physics mentors, as well as to encourage continued and improved support for students with disabilities in the physics community.

II. CLUSTERS OF MODELS OF DISABILITY

To capture complex views of disability in physics mentors’ reasoning for providing specific supports to disabled students, we applied a three-dimensional framework for considering “clusters” of disability models, where models differ in how the cause, effect, and dis/ability dichotomy are construed [10].

A. Cause dimension: social versus medical

The cause dimension describes the underlying cause of disability and provides a distinction between the roles of impairment and disability. The social model focuses on the interplay between impairment (i.e., mind and/or body limitation that is made clear in specific settings [11]) and the environment. Disability occurs when an impaired person’s opportunity to fully engage with the environment is lost due to physical or social barriers [10,12]. An example of the social model is the use of accommodations in the classroom is providing extra test time and extended deadlines [13], which reduce barriers in the instructional environment and provide access. On the opposite side of the spectrum, the medical model focuses on the individual and/or the impairment (i.e., body and/or mind dysfunction [11]) as the cause for not having access, and the burden for change is placed on the disabled individual [10,12]. For example, instructors might inappropriately discuss medication use with a student struggling with time during exams [14].

B. Effect dimension: affirmative versus tragedy

The effect dimension defines disability with respect to quality of life and well-being as a result of having an impairment. Under the affirmative model, it is recognized that impairments do not only cause harm, and a disabled person’s well-being may even be enhanced by their impairment [10]. Additionally, impairments are celebrated as an important aspect of diversity [12]. For example, an instructor might recognize an autistic student as someone who quickly interprets patterns and has an attention for detail [15]. On the opposing side of the spectrum, the tragedy model describes a disabled person’s well-being as diminished due to their impairment, and that disabled people might desire to be able-bodied [10]. Someone who espouses the tragedy model of disability might show pity for disabled people or describe successful disabled people as “brave” for overcoming disability [10]. An example of the tragedy model in action is when instructors demonstrate ableism by acting surprised that a student who has performed well in their course requests to use an extra test time accommodation [14].

C. Dis/ability dichotomy dimension: universal versus minority group

The minority model describes a clear distinction between being disabled and non-disabled (i.e., dis/ability dichotomy) [10,12]. For example, an instructor might only allow changes to the course design for students who seek approval through a Disability Services Office, where students often must disclose and document their disability status to benefit from such services [16]. In the universal model, disability describes the inherent variation in peoples’ needs, abilities, and interests. In this model, disability is conceptualized as a spectrum of capabilities, rather than a dichotomy [10]. Under the universal model, an instructor may intentionally design their course with the variety of student abilities in mind using Universal Design for Learning [17].

III. METHODS

A. Interviews with mentors

Five physics faculty from different U.S. postsecondary institutions participated in remote, semi-structured interviews about their experiences mentoring and teaching
students with disabilities in either Spring or Fall 2022. Participants were recruited via disability-specific physics surveys where they indicated interest in participating in an interview (as described in [18-20]) or personal contacts of the authors. Each interview lasted one to two hours, as selected by the participant; participants were provided with an option to participate in a longer interview if they preferred. Participants varied in terms of their age, gender, U.S. nationality status, and disability status. To protect the anonymity of participants, we only share pieces of identity that help with contextualization of the findings, and we refer to participants by pseudonyms.

Questions in the interview protocol were structured to prompt participants to use different models of disability as part of our investigation of how various models manifest in physics settings. Therefore, participants may have been promoted to use specific models when discussing how they support mentees.

B. Data analysis

Using the six models of disability as a priori codes, the first author (C.M.D.) analyzed each interview to identify instances of each model that manifest within the verbatim transcripts of the interviews. Codes along varying axes are not mutually exclusive allowing for the clustering of models. Then, C.M.D. discussed a single interview with D.O. (co-author) until they reached agreement about the implementation of the codes. Next, C.M.D. combed through the interviews for physics-specific examples of mentor decisions to support students with disabilities. C.M.D. then discussed an example from each interview until agreement was reached with co-authors, J.J.C. and E.M.S.. Afterwards, C.M.D. used constant comparison to compare and contrast decisions made by mentors [8]. While we identified multiple types of mentor supports, we will focus this paper on describing research support and career advice given by mentors to disabled students as these topics are not well described in extant physics education literature.

While we present examples expressed by individual physics mentors, our intention is not to critique these individual instructors. Rather, we recognize views expressed by individual mentors as indicative of the academic physics community, and we understand the academic physics community to both be shaped by ableism, (i.e., valuing and accepting some abilities which influences our worldly perspectives [21]).

C. Positionality and language

Research team members experience a variety of impairments, including emotional/mental health, physical/mobility, health, and hearing impairments (access [22] for explanation of these categories). While C.M.D. identifies with several impairments, she does not have a personal preference between person-first or identity-first language. We use a variety of person-first and identity-first language throughout this study 1) to relate our study with language familiar to mentors and 2) to promote inclusivity with how a student or mentor might identify.

IV. FINDINGS

We frame each example of support described by the participants using clusters of models of disability [10]. The examples presented in this paper do not represent an exhaustive list of supports for students with disabilities. For clarity, we delimited our findings for this paper to examples that can be clearly described by the six models of disability. Future work will include a wider range of examples and possible room for critique of the models.

A. Providing access to research environments for students with mobility impairments

Two mentors, Ren and Brad, both tenure-track faculty at research intensive universities, described supports they have used for graduate students with paraplegia. When asked about accommodations made in their research lab, Ren, an experimental lab-based researcher, discussed two main supports for a graduate student with paraplegia. Ren said, “We worked on making the lab more accessible, and, you know, with the wider corridors … plus seek support from the [college that houses Ren’s department] to buy a special, specialized wheelchair that allowed the student to access a machine...” In this excerpt, Ren describes two accommodations, which each are aligned with a different model of disability on the cause axis. We interpret widening the corridors as aligning with the social model because the physical environment, rather than the individual, was modified. This alteration removed a barrier both for the specific student and future users who may need additional space to maneuver, which potentially aligns with the universal model as well. Additionally, Ren provided access for the student to reach lab equipment by purchasing a specialized wheelchair. The specialized wheelchair does not permanently eliminate the barrier for future users by altering the environment, but rather modifies the individual student’s mobility, so we interpret this accommodation as aligned with the medical model. We posit that an advantage of analyzing the supports physics mentors have provided is to propose more durable solutions. In this case, Ren and equipment producers could apply the social model by implementing principles of Universal Design [23] to make the laboratory equipment usable by a wide range of users.

Responding to a similar prompt, Brad, a physics researcher whose research mainly takes place outside of a physical lab, described how they accommodated a student with paraplegia by holding remote meetings using videoconference software. Brad commented, “...we've never had a conversation about the wheelchair per se, it's just a thing that's true and obvious and made me realize that getting to my office is really not accessible for them ... and so a lot of those meetings on Zoom for other reasons, but a benefit of
having them on Zoom is that the student doesn't have to worry about things like getting into my office or getting to places that are maybe not easy to get to.” Like Ren, Brad’s support aligns with the social model since an environmental barrier to access and participation was eliminated. Brad’s accommodation is aligned with the minority model by initially using the accommodation for a single student. Later, Brad mentioned using the same accommodation for colleagues located in other places, perhaps due to the increased accessibility of technology as a result of the COVID-19 pandemic [24]. Brad noted, “And that's accessible in the sense that people can be wherever they want and whatever is comfortable for them.” This perspective might align with a common critique of the universal model: interest convergence. Interest convergence refers to the idea that equality occurs when interests of a minoritized group are in alignment with the interests of the majority group [25]. Within this context, it is possible that giving equity-creating support to every person undermines the needs of people with disabilities who originally needed that support to have equal opportunities for access. Further research is necessary to examine the effects of interest convergence on disabled students’ participation in physics spaces.

Ana is a senior professor at a small private university with the leading role of teaching physics courses. During the interview, Ana described using her van to assist mobility-impaired students with transportation to research conferences. Ana explained, “… I have a van which I use for [family member] with the tie down things [straps], so I can do that to transport the students, but actually, even sometimes uh, teachers sort of borrowed my van to do that, because not everybody have their own van to transport that. Actually recently, when I complained, uh University purchased the minivan with the wheelchair accommodation.” After complaining to the university, Ana successfully removed the barrier for the students with disabilities, which is aligned with social model. Initially, acquiring a van after emphasizing the need for help with transporting disabled students to conferences aligns with the minority model since the focus is supporting students with mobility impairments. Afterward, the school’s resources now provide access for a wider range of users, which is in line with the universal model. In Ana’s example, accommodations for disabled students might not happen unless faculty create pressure to make change.

B. Supporting students with mental health impairments

Will, Brad and Ren all discussed examples of providing advice to disabled students about their future career or continued involvement with physics. Will teaches physics courses at a two-year college, and Brad and Ren both mentor physics students at a research-intensive university.

When asked about the impact accommodations have on preparing students for their career, Will described a conversation with a student with anxiety who used an extra time accommodation for class exams: “I said, "Well, you know, what is it you want to do?" Wants to be an ER physician. And I'm sitting there thinking, "Wait a minute, you have extreme anxiety under stress, and you want to be an ER physician? Um, maybe someone needs to have a conversation with you.” He's like, "Oh, no, I've already looked into it. You know, the medical schools will accommodate that." I'm just going, "Wait a minute. I don't want you as my ER physician that, 'Okay, give me ten minutes. He may be spouting blood; I need ten minutes here'." In their conversation with the student, Will fixed on the student’s impairment inhibiting their ability to treat a patient, which closely aligns with the tragedy model. Since lower levels of anxiety positively impact student performance [26], Will could potentially better support the student by considering benefits of anxiety that may enhance a doctor’s practice like preparing multiple plans of action [27].

Like Will, Brad described a time when they encouraged a student with mental health impairments to re-consider their future career, specifically to leave the physics graduate program. Brad said, “Fundamentally, what I, what I felt was this student needed the support of successfully finishing something, while also getting away from the toxic environment that grad school can be, which was not helping them. And so, I supported that, I supported them in doing that ... They eventually left with [a] masters, and I hope they're doing better.” Brad points out that the student’s impairment and the toxic environment of graduate school are both harming the student’s well-being, which is in line with the tragedy model. However, since the mentor fixed on removing the student from the environment rather than mending the environment, Brad’s support aligns with the medical model. From Brad’s example, it is possible some mentors might need help supporting students with disabilities because the scope of support is beyond what a mentor can change on their own, which might prompt a mentor to encourage a student to leave a program. Instead, physics graduate program coordinators might consider modifying departmental policies to create a more supportive environment (i.e., reform key components of graduate program like professional development, advising requirements, curriculum and candidacy exam [28]).

Ren shared the idea that internships outside of academia might help a student with anxiety, especially anxiety exacerbated by imposter syndrome (i.e., feelings of inadequacy pertaining to ability with role performance [29]). Ren responded to a prompt about how a student’s impairment positively impacted their physics experience, by commenting, “The positive aspect, I think the student asks really sharp questions. Um what's surprising is that when I say, ‘that, that was really good questions that you're asked,’ it doesn't make this person feel better.” Here, Ren’s realization that the student’s questioning skills are positively impacted by the student’s anxiety is in line with the affirmative model. While Ren states that these praises do not seem to make the student feel better, implying that Ren
think the student’s impairment is negatively impacting their well-being, we still interpret this excerpt as aligned with the affirmative model, since the tragedy model only allows for a negative perspective of disability. Ren continued to discuss their plan of support, which involved sending the student to an internship outside of academia. Ren said, “...what I’m trying to do is to basically send this person to internship where they can actually work with professional scientists that work regular job, unlike ourselves as a physics professor, so that they understand and then talk to real people.” Here, Ren’s plan is embedded in the social model because Ren is changing the environment to support the students’ well-being and career success. While Ren’s support plan aligns within the minority model due to its implementation with a single student, Ren’s perspective of internships holds a universal model approach as they continue to provide more details. Ren continued, “I think everyone should be sent to internships anyways. Yeah, because a lot of students who do so well in undergraduate physics program, they might come into grad school thinking that they should just do what they’re best at, and then they might not know exactly what they want to do after graduating from grad school.” In Ren’s excerpt, the idea that students can benefit from an internship outside of academia lies within the universal model because there is a potential to reduce imposter syndrome and provide useful resources to all students. Additionally, Ren’s support is an example of a resource that can counter the narrative painted by Sophia below related to non-inclusive physics culture. In general, postsecondary programs can leverage resources like internships in the STEM community to support students with disabilities and historically underrepresented groups [30].

C. Preparing students for physics culture

Unlike other mentors, Sophia, who teaches physics at a two-year college, talked about conversations with students about the culture of physics rather than discouraging students with disabilities from pursuing their desired career. In their interview, Sophia mentions the unwelcoming culture of physics to be a barrier for people with disabilities, and maybe even specific disabilities. Sophia stated, “I think there’s a lot [of barriers] in physics as a culture. But it does seem to me like there, it’s still very much focused on being, you know, perfectly able-bodied and be brilliant in certain ways. ... And when I think, you know, it’s specifically for people with disabilities that is not welcoming and not inclusive and not um yes, just not okay.” Furthermore, Sophia warns students from minoritized groups about the unwelcoming culture of physics. Sophia commented, “So I don’t think I’ve ever discouraged or at least not consciously discouraged someone because of an impairment or disability. Um, I can imagine having conversations with a student who is really interested in physics, but was, well and I have had, they were a minority, they’re just minorities in other ways. You know, the reality is the [physics] culture tends to be fairly unforgiving of differences.” We interpret Sophia’s advice as aligned with the minority model, since she uses the word “minority” and only provides a warning about the culture of physics to specific students. Both of Sophia’s excerpts are possibly in line with the tragedy model because of the implication that minoritized groups, including students with disabilities, struggle with fitting in to physics spaces, which might impact a student’s well-being. Instructors could consider discussing the physics culture with the whole class, including brainstorming ideas about how the instructor and class members could create a more inclusive environment within and beyond the classroom.

V. DISCUSSION

In this paper, we provide examples of how mentor supports map onto clusters of models of disability in physics settings. The affirmative model was the least prevalent model used by mentors when supporting students in this data set. Future work will investigate the prevalence of each model and identify opportunities for mentors to use the affirmative model.

We do not pinpoint a cluster of models that is the most appropriate to use in academic contexts. Rather, we provide examples of support linked with models of disability with the intention of sparking reflection within the physics community on what supporting disabled students currently looks like and potential room for improvement. For example, the social plus universal cluster of models aligns well with implementing Universal Design for Learning-aligned instructional practices. The elimination of barriers that deny some students access to course material fits within the social model while implementing practices that support access for all students without having to make an accommodation for individuals fits within the universal model. By using such practices in classrooms and research settings, mentors and instructors potentially eliminate the need to make retrofits to their design while considering inclusion.

VII. LIMITATIONS AND FUTURE WORK

The main limitations of our study are our sample size and sample demographics. Our sample does not include mentors in industry or races and ethnicities historically underrepresented in STEM. In future work, we will continue to recruit participants from a variety of backgrounds. However, it is possible individuals from minoritized groups in mentor positions already feel burdened to share their experiences [31,32]. Also, to solidify the effectiveness of the support provided by the mentors, we will triangulate our findings from this study with interviews with mentees and we will follow-up with participants (i.e., member checking).

ACKNOWLEDGEMENTS

This work is supported by NSF CAREER Grant 1750515.
[31] L. E. Hirshfield and T. D. Joseph, ‘We need a woman, we need somebody new to do it?’ Race and cultural taxation in the academy, Gender and Education, 24(2), 213-227 (2012).
Students’ sense of belonging predicts their success and persistence in STEM courses. Collaborative, small-group activities form the foundation of many research-based instructional strategies. Our broader project seeks to understand the role of small groups in students’ sense of belonging to support instructors in the formation of equitable groups in active engagement classrooms. In this article, we focus on the construct and discrimination validity of a belonging measure. To assess the belonging measure’s ability to discriminate across time, courses, and demographic groups, we administered a short survey on belonging in a variety of STEM courses that used groups as a pre- and post-class assessment. We analyzed the results using structural equation modeling to inform the validity of the survey and identify possible differences of interest. The results provided evidence for both construct and discrimination validity. Belonging varied across the courses and changed from pre to post in two of the four courses: one course saw a decrease and the other course saw an increase. Men tended to have a higher sense of belonging than women and the changes in belonging increased these gender differences. One possibility is that the differences observed across courses could result from the different practices used to support group work within each course. The validity evidence for the belonging measure indicates it will support our ongoing research to establish the statistical relationships between instructor practices to implement and support small groups and students’ sense of belonging.
I. INTRODUCTION

Student belonging in the classroom affects persistence in STEM, especially for women and minoritized individuals [1, 2]. In particular, past research links student belonging to higher self-efficacy and motivation [3, 4], more engagement [5], and increased persistence [6–9]. Since not all students feel like they belong in STEM [10–14], some instructors have incorporated interventions designed to support student belonging, often to good effect [15, 16]. Many such interventions use student groups to leverage the connection between peer groups and sense of belonging [8, 14, 17, 18].

Student-centered, collaborative instruction can lead to more learning and higher grades than lecture-based instruction [19–22]. Such instruction may [23, 24] or may not [25] achieve improved equity. These practices often have students work in groups of three to five students. Best practices for small group work are not well established. For example, some studies support the formation of groups with heterogeneous prior achievement [24, 26–28], others argue for creating groups with homogeneous prior achievement [29–32], and a few suggest it does not matter [33–35]. Another common grouping strategy is the avoidance of solo status for minoritized students [36]. Laboratory research and stereotype threat [37–39] both indicate the harm that solo status can cause. We are not, however, aware of any studies in the classroom that indicate this harm occurs or the size of the harm.

Our overall project investigates the effectiveness of various grouping strategies instructors can use to form intentional groups that contribute to student outcomes and classroom equity. In this article we focus on measuring students’ sense of belonging in their class. While our broader project investigate a variety of student outcomes including self-efficacy, conceptual understanding, and course grade, this article focuses on substantiating the construct and demonstrates evidence for construct validity.

II. THEORETICAL FRAMEWORK

We use Connell & Wellborn’s self-system model of motivational development, which they based on the fundamental human needs for competence, autonomy, and belonging [40]. This model supposes that the classroom impacts students’ self-perception, including their sense of belonging, which in turn impacts their engagement and achievement [3, 40–42]. We further draw on an anti-deficit approach with the perspective that students enter a course with abilities, talents, and capital, and that the institution is then responsible for structuring a course to support all students [43–46].

III. METHODS

We studied instructional practice and students’ sense of belonging in four courses. In each course, instructors randomly assigned groups of 3-5 students. All courses used frequent active learning in the classroom and encouraged student-student interactions around content. We surveyed students during the first and last 2 weeks of the course. Table I summarizes group work practices in each course, including the use of undergraduate learning assistants (LAs), who support instructors by co-facilitating small group work [48]. The use of LAs in the classroom has been shown to strengthen students’ sense of belonging in STEM [49, 50].

Table II gives the course demographics for students who consented to participate. Data collection included many social identity groups, which we grouped in the table due to small sample size. The minoritized group included Black, Hispanic, Indigenous, Pacific Islander, and Middle Eastern students. We received responses from non-binary and trans students, but we were unable to perform statistical analysis for these groups due to small sample size. The demographics for each class were typical for the corresponding field.

Course A was algebra-based physics I, mainly taken by students in their final two years, with 690 health-related students at a large 4-year public university that is an emerging Hispanic-serving Institution (HSI). Each week the course met for two 50-minute lectures—including conceptual questions, worked examples, demonstrations, and think-pair-share iClicker questions—and two discussion section meetings with a TA—including homework help, quizzes, and group work. Students received participation points for working together (during about 25% of contact time) in assigned groups of 5 (with 6 groups per TA) on worksheets solving problems.

Course B was a core, upper-level biology course for majors with 62 students at a small, private, 4-year liberal-arts HSI. The course consisted of three 55-minute lectures and a three-hour lab each week. The instructor lectured with frequent short discussions, with about 20% of contact time spent on group work, and there was an out of class group assignment.

| Table I. Instructional practices in each course [47]. |
|-----------------|-----------|-----------|-----------|-----------|
| Course A B C D |
| Learning Assistants (LAs) support group work       | -         | X         | X         | X         |
| Groups discuss expectations and draft contract    | -         | X         | -         | -         |
| Groups work on structured assignments in class    | X         | X         | X         | X         |
| Group members assigned changing roles             | -         | -         | -         | X         |
| Groups work together outside of class             | X         | X         | X         | -         |
| Peer evaluation                                   | -         | X         | X         | -         |

89
Course C was a sophomore-level engineering course with sections of about 30 students each (177 total students), at a mid-sized, 4-year, public, emerging HSI. Classes met three times a week for 100 minutes. Instructors presented the topic through an interactive lesson on the whiteboard with example problems. Students used 50% of contact time to work on problems at their tables in groups of about 4 as the instructor and one to two LAs circulated to facilitate group work and engage students in discussion regarding their problem-solving. There were about 3-4 groups per LA. In addition to these ungraded group activities, students worked on a graded, out-of-class project.

Course D was the first course in the general chemistry series for life science majors at a 4-year, research-intensive, public, emerging HSI. The course was taught by an instructor, four TAs, and approximately 20 LAs, with around 130 students enrolled in a section (266 total students). The course consisted of three 50-minute lectures and a 110-minute discussion each week during a ten-week quarter. The lecture consisted of three 50-minute lectures and a 110-minute discussion each week during a ten-week quarter. The lecture used think-pair-share and clickers. The discussion section was based on a blended model of Process Oriented Guided Inquiry (POGIL) and Peer-Led Team Learning (PLTL). Students worked in small groups on structured worksheets designed around the learning cycle. LAs supported about two groups each. Teams also completed the second stage of the two midterms together.

We used a six-item instrument to measure social belonging on a six-point Likert scale (strongly disagree, mostly disagree, somewhat disagree, somewhat agree, mostly agree, strongly agree), drawn from Fink et al. [10]. The instrument (see Table III) includes two factors: perceived belonging (first four items) and belonging uncertainty (last two items). Perceived belonging relates to the general feeling of belonging in a course, while belonging uncertainty relates to the stability of that belonging. We administered the pre- and post-course belonging surveys through the LASSO Platform [51] for Courses A, B, and C; course D administered the surveys through their learning management system.

We used structural equation modeling (SEM) [52] to investigate the three research questions. For RQ1, we used confirmatory factor analysis (CFA) [52] to test the construct validity of the belonging measures for our study population. We first tested the model based on Fink et al. [10] using the lavaan [53] package. When necessary to improve model fit, we used the modindices command. We repeated these steps until the factor loadings and the fit indices passed the cutoffs discussed in the next paragraph.

The root mean square error of approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis index (TLI), and factor loadings informed how well the model fit the data. RMSEA is an absolute fit index that addresses parsimony in the model by accounting for the model’s degrees of freedom. High RMSEA indicates an over-constrained model with too few degrees of freedom. RMSEA has several proposed cutoffs, < 0.05, 0.06, 0.08, and 0.10, indicating good to acceptable fit [54]. CFI and TLI are relative fit indices. They compared the test model to a baseline model with all covariances set to zero and all variances freely estimated. Both indices range from 0, no fit, to 1, best possible fit. We used a cutoff of CFI and TLI > 0.95 [54]. Balancing absolute and relative fit indices can lead to a model that is simple and fits well. We report scaled fit indices. Hair et al. [55] proposes factor loadings of 0.5, explaining 25% of the variance in the item, as an absolute minimum and 0.7 as a preferred minimum, explaining 50% of the variance.

To test RQ2 and RQ3, we used multiple group SEM. This method built separate SEM models for each course (i.e., multiple groups) to investigate the shift from pre- to post-instruction. To determine if differences existed across courses, we compared these SEMs to a set of SEMs that constrained the intercepts and regression coefficients to the same value across all of the courses. We used the RMSEA, CFI, and TLI to assess the model fit. To determine if the unconstrained multigroup SEMs provided unique information, we compared them to the constrained multigroup SEMs with an ANOVA. We set the intercept for the models to the largest

<table>
<thead>
<tr>
<th>Question</th>
<th>Factor Loading Pre</th>
<th>Factor Loading Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>I feel like I fit in this course.</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>Setting aside my performance, I feel like I belong in this course.</td>
<td>0.9</td>
<td>1.0</td>
</tr>
<tr>
<td>I feel comfortable with my peers and classmates in this course.</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>I feel comfortable with my instructor(s) in this course.</td>
<td>0.7</td>
<td>0.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question</th>
<th>Factor Loading Pre</th>
<th>Factor Loading Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>I feel uncertain about my belonging in this course.</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>If I don’t perform, I feel like maybe I don’t belong in this course.</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>
FIG. 1. Belonging across the four institutions constructed from the averages of the questions in the two factors, which are shown in Table III. The boxplots show the distribution of the data with outliers as large black dots. The light grey dots represent individual students with a slight jitter to spread out the data.

Regression coefficients represent differences from the intercept in units of standard deviation (SD). The SEM builds the latent variables as normal continuous distributions. The SEM calculates thresholds for the point on this continuous distribution that a response would shift (e.g., going from strongly agree to agree) such that our six-response Likert-scale had five thresholds. The thresholds tended to cover a span of 3 SD, which means going from strongly disagree to strongly agree is approximately a 3 SD shift. Because of this scale and the spread of the results in the data, we adopted Cohen’s [56] rules of thumb for effect sizes:

- $< 0.2$ is very small,
- $0.2$ to $0.4$ is small,
- $0.4$ to $0.8$ is moderate, and
- $> 0.8$ is large.

IV. RESULTS

Figure 1 presents the perceived belonging and belonging uncertainty (pre and post) for the four courses constructed by averaging the questions for each factor. Perceived belonging tended towards ‘mostly agree’ in all courses. Perceived belonging was lower in Course A and highest on the posttest in Course D. The only notable shift in perceived belonging was for Course D where both the median and inter-quartile range increased. Belonging uncertainty tended to be between ‘slightly agree’ and ‘slightly disagree’. Consistent with perceived belonging, the noticeable shift in belonging uncertainty was for Course D where both the median and inter-quartile range shifted down, indicating higher belonging.

We tested the two-factor structure of the sense of belonging scale using CFA for the pretest and posttest data. All indices indicated a good fit: $\text{CFI}_{\text{pre}} = 0.997$, $\text{CFI}_{\text{post}} = 0.978$, $\text{TLI}_{\text{pre}} = 0.982$, and $\text{TLI}_{\text{post}} = 0.953$ ($> 0.950$ indicates good fit), $\text{RMSEA}_{\text{pre}} = 0.039$ and $\text{RMSEA}_{\text{post}} = 0.059$ ($< 0.060$ indicates good fit). The factor loadings, shown in Table III, indicate adequate fits for all items with the lowest fit for the question about performance for the belonging uncertainty factor.

We investigated the change over time in the courses (RQ2) using multi-group SEM, and identified several relationships that also appear in Fig. 1 of the average scores. We did not include Course B in the analysis because no students responded to ‘strongly disagree’ or ‘disagree’ for several questions on belonging. The ANOVA between the constrained and unconstrained SEMs was statistically significant ($p < 0.001$), indicating variation across the courses. The fit indices for the unconstrained SEM indicated adequate fit: $\text{TLI} = 1.00$, $\text{CFI} = 1.00$, $\text{RMSEA} = 0.096$. The regression coefficients in Table IV indicated small to moderate improvements in Course D, little to no change in Course C and very small decreases in belonging in Course A.

Due to small sample sizes, we restricted our analysis for RQ3 to investigate gender differences in Course A and D. The multigroup (course) SEM created standardized regression coefficients, shown in Fig. 2 and Table IV, for men’s pretest and posttest scores and women’s posttest scores. The intercept was women’s pretest scores. In both courses, men tended to report a higher belonging and lower belonging uncertainty than women both before and after instruction, in agreement with previous findings [11–13, 57, 58].

In course A, students’ perceived belonging decreased and belonging uncertainty increased. The decrease in perceived belonging for men shifted moderately (0.3 SD) and was statistically significant. The other shifts, while all being consistent with this decrease, were much smaller. The shifts in Course D were also consistent in that perceived belonging increased and belonging uncertainty decreased, and the shifts and differences across gender were much larger than in Course A.

### TABLE IV. SEM regression coefficients for both models. The intercept was set to either the pretest or the pretest for women.

<table>
<thead>
<tr>
<th>Course</th>
<th>Time</th>
<th>Gender</th>
<th>Belonging Coef.</th>
<th>SE</th>
<th>Belonging Uncertainty Coef.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>Pre</td>
<td>-</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>A</td>
<td>Post</td>
<td>-</td>
<td>-0.181</td>
<td>0.072</td>
<td>0.122</td>
<td>0.069</td>
</tr>
<tr>
<td>C</td>
<td>Post</td>
<td>-</td>
<td>-0.149</td>
<td>0.154</td>
<td>0.003</td>
<td>0.114</td>
</tr>
<tr>
<td>D</td>
<td>Post</td>
<td>-</td>
<td>0.663</td>
<td>0.098</td>
<td>-0.314</td>
<td>0.097</td>
</tr>
<tr>
<td>NA</td>
<td>Pre</td>
<td>Women</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>A</td>
<td>Pre</td>
<td>Men</td>
<td>0.452</td>
<td>0.118</td>
<td>-0.316</td>
<td>0.101</td>
</tr>
<tr>
<td>A</td>
<td>Post</td>
<td>Women</td>
<td>-0.134</td>
<td>0.088</td>
<td>0.110</td>
<td>0.082</td>
</tr>
<tr>
<td>A</td>
<td>Post</td>
<td>Men</td>
<td>0.149</td>
<td>0.120</td>
<td>-0.269</td>
<td>0.108</td>
</tr>
<tr>
<td>D</td>
<td>Pre</td>
<td>Men</td>
<td>0.181</td>
<td>0.140</td>
<td>-0.323</td>
<td>0.147</td>
</tr>
<tr>
<td>D</td>
<td>Post</td>
<td>Women</td>
<td>0.574</td>
<td>0.119</td>
<td>-0.217</td>
<td>0.120</td>
</tr>
<tr>
<td>D</td>
<td>Post</td>
<td>Men</td>
<td>1.101</td>
<td>0.168</td>
<td>-0.217</td>
<td>0.120</td>
</tr>
</tbody>
</table>
Women started (0.3 SD) and ended (0.5 SD) with higher belonging uncertainty than men and belonging uncertainty decreased for both men (0.4 SD) and women (0.2 SD). Women started (0.2 SD) and ended (0.5 SD) with lower perceived belonging than men and perceived belonging increased for both women (0.6 SD) and men (0.9 SD). Both the improvements in belonging and the greater improvements for men were consistent across both perceived belonging and belonging uncertainty. The shifts from pre to post were small to large in size. While belonging improved in Course D, these improvements were larger for the men in the course than for the women.

V. DISCUSSION AND CONCLUSIONS

The CFA provided evidence for the construct validity of the belonging survey. All of the indicators of model fit met or exceeded the minimum values. The belonging measure also captured differences across courses, shifts in belonging for groups of students, and differences across demographic groups. The two latent factors, perceived belonging and belonging uncertainty, had very different average responses. Perceived belonging was mostly positive across courses, with students tending to ‘mostly agree’ that they belonged in the course. Responses to belonging uncertainty tended towards ‘slightly disagree’ with a larger spread in all of the courses than the perceived belonging responses. The changes also varied across the courses, but were consistent for both belonging and belonging uncertainty. In courses B and C, there was little shift in either perceived belonging or belonging uncertainty. A small negative shift occurred in course A. The largest increases occurred in Course D with an increase in perceived belonging and a decrease in belonging uncertainty.

The most noteworthy shifts in student belonging occurred in course D, in which perceived belonging increased and belonging uncertainty decreased for both men and women. On the other hand, course A experienced shifts in the opposite direction, though they were not as large. We suspect the use of instructional practices oriented toward supporting group work (shown in Fig. 1) may contribute to the students’ sense of belonging in these courses, especially given that course A made only limited use of structured group assignments in class. Prior research suggests that one of these practices, the use of LAs, may be particularly effective at improving students’ sense of belonging [49, 50]. However, course C experienced at most small shifts, despite use of LAs. Because there are many other differences between the courses, including institution and field of study, we hope that expanding the courses in our dataset will help identify the factors contributing to belonging.

The gender differences present in the chemistry course (D) and not the physics course (A) stand out from prior work showing larger gender differences on many affective outcomes in physics than in chemistry [59, 60]. This finding also contrasts recent work showing that rotating roles can benefit women in particular [61, 62], as only Course D used this practice. While these results indicate discrimination validity in the belonging measure, the small sample in this study limits our ability to investigate the causes of these differences. The results do point to the need for the larger study across the intersections of discipline, gender, and race/ethnicity that we are pursuing.

These findings support our use of this belonging measure in our larger study of instructor practices to support group work. The construct and discrimination validity indicate the instrument can identify differences across courses and likely could identify differences between groups or students randomly assigned to different conditions. Ongoing data collection will support several investigations. Larger samples will allow modeling the relationships between the instructor practices in Table I and student belonging, which we can only speculate upon here based on prior research. Ongoing research will also look at differences in belonging and learning outcomes for groups. We will investigate if groups with either homogeneous or heterogeneous prior performance lead to different student outcomes and outcomes for students in solo-status conditions due to either their race or gender.


[34] F. Lawrenz and T. L. Strayhorn, Mentoring and Satisfaction with College for Black Students, CBE - Life Sciences Education 12, 010107 (2023).


[54] Y. Xia and Y. Yang, RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods, Behavior research methods 51, 409 (2019).


[58] D. R. Johnson, Women of color in science, technology, engineering, and mathematics (stem), New Directions for Institutional Research 2011, 75 (2011).


Learning assistants’ teaching strategies for promoting scientific inquiry among undergraduate students in a physics laboratory setting

Samuel W. Engblom, Maggie S. Mahmood, Michael J. Matos, Jackie Vargas

Physics Department, University of Illinois Urbana-Champaign, 1110 W. Green St, Urbana, IL, 61801

Introductory labs in STEM courses are often designed with the goal of promoting scientific reasoning among students. Instructors play a role in reaching this goal, and the pedagogical practices instructors adopt can support or inhibit the development of student scientific reasoning. In particular, novice instructors who are still grappling with lab epistemology may enact a range of teaching practices, some of which may not be helpful in reaching the goals of lab instruction. We present findings from a study of 58 first-time undergraduate Learning Assistants (LAs) presented with a teaching scenario from an introductory physics lab. In our analysis, we identify teaching strategies used by LAs to support the scientific reasoning of their students and barriers or alternate approaches that prevent LAs from successfully providing this support. We find that a common barrier to LA support of student scientific inquiry is LAs’ preconceived notions of a “correct answer” to the lab task that students must reach, which causes these LAs to shift to strategies more in line with conducting validation experiments.
I. INTRODUCTION

Physics labs have been the subject of substantial research and reform in physics education. Pedagogically, many lab reforms have aimed to make the practices that occur in physics labs mirror the disciplinary practices of physics in an effort to promote student scientific thinking and lab skills [1–4]. In moving toward the goal of creating labs that meaningfully impact student scientific thinking, two critical features have emerged from research: labs must allow students to participate in decision-making and do away entirely with goals that ask students to verify established theories [5]. The latter of these two features can be challenging to implement, since many students believe the aim of introductory labs is the verification of physics theory [6, 7], and will readily shift toward framing labs as validation experiments if given any sort of cue or hint to do so [8].

Reformed labs have benefited from increased instructional support in the form of additional instructors present to assist students. One potential source of additional instructional support has been near-peer instructional programs. The Learning Assistant Program is one such model for near-peer instruction in which undergraduate students — often former students of the course in question — act as co-instructors in a course [9]. The presence of undergraduate Learning Assistants (LAs) in physics learning environments has been connected to lower failure rates [10–12] and gains on physics concept inventories among students [13–15]. A central tenet of the Learning Assistant model of near-peer instruction is that LAs are exposed to discipline-based pedagogy and research through a pedagogy course in which every first-time LA must enroll [9]. Studies of LA pedagogical knowledge development have shown that LAs more readily take up some pedagogical ideas than others [16]. Beyond any formal training they may receive, LAs also bring in numerous ideas about what teaching and learning should look like as they navigate the boundary space of being both students and teachers [17].

Compared to research on the practices and outcomes of LA programs in discussion and recitation sections, the results of their implementation in physics labs have been underreported, though LA programs have been successfully implemented and studied in lab settings within other STEM disciplines [18–20]. Lab instruction and pedagogy can be messy — students often hold conflicting and seemingly contradictory epistemological beliefs that differ from the beliefs of their instructors about the purpose of labs [6, 7]. LAs, being experienced students but relatively new teachers, are likely to carry some of these conflicting beliefs as they draw upon their own experiences in their teaching, even when a “preferred” pedagogical approach is presented as part of the pedagogy course and weekly lab preparation meetings.

We seek to study the pedagogical practices of LAs in introductory physics labs and how they align or misalign with the goal of helping their students develop scientific reasoning skills. To this end, we have analyzed 58 first-time physics lab LA responses to a teaching scenario presented as a free response question in the LA pedagogy course. In our study, we have considered the following two research questions:

1. What pedagogical practices do lab LAs take up that support student scientific reasoning?
2. What barriers keep LAs from taking up such practices?

II. METHODS

In our analysis, we used an inductive coding approach to identify and code themes within a set of LA written responses [21]. Our coding approach focused on how LAs centered data or physics concepts in their responses, and statements LAs made about their hypothetical teaching moves. A second iteration expanded the scheme to codify specific teaching moves, as well as whether LAs indicated within their explanation whether they believe the task presented in the scenario had a single correct answer.

A. Study Context

The data for this study were taken from a free response question posed to LAs as part of the LA pedagogy course in the Fall 2022 and Spring 2023 semesters at a large Midwestern university. All the scenario responses were from first-time LAs. The pedagogy course included LAs from both introductory mechanics and introductory electricity and magnetism lab courses. Each semester, LAs are recruited from former students of the course — a noteworthy detail, as it means that each LA has previously been a student in the course in which they instruct, often as recently as one semester before. In the first weeks of the pedagogy course, LAs are presented with concepts such as open and closed questioning and science talk moves [24] to help them establish a baseline set of strategies for instruction.

In our context of study, students in the labs use the Interactive Online Laboratory tool (IOLab) to complete laboratory tasks. The IOLab system includes a multi-sensor measurement device and data analysis software that receives and displays data from the device [22]. In labs, the students are given ill-structured tasks designed to encourage sense-making and build adaptive expertise [3, 23]. These lab sections are generally staffed by one graduate teaching assistant and 1–2 undergraduate LAs.

We identified this set of responses for analysis due to its richness and variety, and the emergence of two divergent themes with respect to LAs’ lab instructional approaches. When presented with a hypothetical teaching scenario from one of the labs, some LAs focused on encouraging students to use the data to make a claim, while others prompted students to think about physics concepts to make sense of the data. We also saw LAs reference a range of teaching moves. Some LAs specifically called out their use of the strategies from the pedagogy course, while others cited different teaching strategies they had learned through their own experiences.
B. The Scenario

In the prompt, LAs were presented with a scenario situated within a lab assignment from their introductory mechanics course, in which students were asked to measure the acceleration of their IOLab device as it rolled up and down a ramp. Students were asked to test the claim that "the IOLab’s acceleration on the ramp is the same on the way up as it is on the way down." In the scenario, LAs were told that four students were arguing over the results of their data, with two students claiming the accelerations were the same while two other students claimed the up-acceleration was larger than the down-acceleration. The scenario ended by asking the LA how they would respond, with a final clarifying statement that their response would be graded on completion and not on any assessment of the correctness of their answer.

In the actual lab activity the scenario replicated, students performing the lab could either accept or reject the claim that the accelerations were the same depending on their data. For some experimental setups, the friction in the wheel bearing could be sufficient to make the accelerations differ when compared via a statistical t' calculation [25]. For other setups, the error analysis may not be able to conclusively show any difference between the accelerations. The goal of the activity was therefore not that students arrive at a "correct" answer, but that students make a claim grounded in their data.

In our coding, we aimed to deduce each LA’s perception of whether or not there exists a "correct answer" to the students’ problem within the scenario, as a glimpse into each LA's epistemological views about labs. Since the LAs were prior students of the course, each of them had performed the lab activity described in the scenario, and may have recalled the answer they obtained when they did the experiment, or generated ideas about what they thought the correct answer should have been by considering physics principles. Ideally, LAs should not pick a side, but should instead encourage students to make their conclusions based on data, rather than based on what the LA or students expect to happen.

C. Analysis

The responses were first read individually by the authors and loosely coded for statements in which the LA focused the students’ attention on either relevant physics concepts or on the data the students had taken. After this first open round of coding, we discussed themes noted in each response, and devised a set of codes to capture how LAs used appeals to data and to physics concepts in their response — ultimately referred to as "pathways" with responses coded as data-driven, theory-confirming, or neither. We later refined the coding scheme to include a new set of codes for the types of teaching moves LAs alluded to within their responses. Based on our emerging sense that LAs’ epistemology could underlie their instructional approaches, we added codes for if the LA's response made any mention of whether the lab the students were working on had one correct answer. As the codes representing epistemology were binary and constrained to each LA’s singular response in time, they served as a means of linking each LA’s personal understanding of the supposed outcome of this specific lab to their instructional approach.

At each step of coding, two independent coders worked together to generate codes for the first 20 responses to establish agreement, then coded the remaining 38 responses separately to allow for the calculation of inter-rater reliability (IRR). In each case, IRR was calculated using Krippendorff’s α [26]. For the pathways, IRR was α = 0.961. For the coding of whether or not the LA believed the lab task possessed a single correct answer, IRR was α = 0.801.

III. RESULTS

In this study, LAs took one of three pathways to addressing the lab scenario outlined in the prompt: (A) Guiding students to focus on and/or discuss their data within their group, then asking them to consider valid scientific interpretations of these results (classed "data-driven"; n = 22); (B) Guiding students to focus on the physics disciplinary content underlying the scenario and use it as a lens for considering the validity of their data (classed "theory-confirming"; n = 15); and (C) Avoiding mention of specific guidance, instead focusing on supporting better communication among group members in their response (classed "null"; n = 21). Our analysis seeks to unpack the pedagogical strategies LAs implement in guiding students through the inquiry process in a manner consistent with the scientific process, while considering what epistemological factors or teaching beliefs may underlie LA decisions to take an alternate path.

A. Teaching Moves Associated with Data-Driven Inquiry

In addressing research question 1, we identify some teaching moves associated with the pathway aligned with an authentic experimental design process wherein data drives the conclusion (pathway "A", "data-driven"), counter to a pathway that seeks to bring data into alignment with a pre-established physics outcome (pathway "B", "theory-confirming").

Opening with direct instruction. The majority of LAs who took pathway "A" gave direct instruction or asked very direct, closed questions at the beginning of their response. For example, in the following response, the LA asked the students to calculate a statistical t’ and use this value as the basis for addressing the physics claim:

I would tell them to look back at the statistical methods provided. It is a tenet of science that what the results say do not depend on the one reading them. If there [sic] are all looking at the same data, if they use the t prime value they
should be able to judge whether it is slowing down on the way down or not.

Here, the LA valued a set procedure for determining whether the data set representing the cart’s acceleration up the ramp was statistically significantly different than the data set representing the cart’s acceleration down the ramp.

In some cases, LAs taking pathway “A” interpreted student conflict as rooted in students’ inherent mistrust of the data.

I would first ask to see their IOLab data and ask them what their conclusions are based off the data. Then I would ask why there is a disagreement. After that, I would say if there is an assumption that your data is incorrect, then address possible errors and what you think the correct answer is.

In this case, the LAs encouragement of the group to refer to the data and allow that to drive their conclusion was rather direct, indicating staunch alignment with an interest in centering the data to drive student conclusions. Conflicts in the group related to physics disciplinary understanding and theories about how the data “should” look, were not attended to.

Establishing student “trust” in the data. In five of the responses, all aligned with the data-driven pathway (“A”), LAs approached the challenge of student doubt by asking the students to retake their experimental trials more carefully, as to improve student trust in the data to drive their eventual conclusions. For example, one LA writes:

If time permits, it might be useful to repeat the trials and see if this is an experimental error. If not, I would first ask them to review their data and see how consistent their result is. Else, it is totally okay to have different opinions. I would ask students to discuss the opposing opinions in their discussion session on their lab report and elaborate on the reasons backing up each claim.

This LA made two distinct moves to establish trust in the data. First, they suggested additional trials, to see if the results in a second set of trials would be consistent with the first. Second, the LA also deemed it appropriate for members of the group to disagree, and take sides in the lab write-up—a move that explicitly valued differing interpretations and reinforced that there was no single correct answer.

Another LA took a similar approach to dispelling student concerns about data, orienting the conversation around specific sources of experimental error:

I would ask them to look over the data that they have. Whether or not the measured acceleration of both the upwards and downwards time frames were the same or not, I would ask them if they think that there was any error that could cause the two accelerations to look different. Since the data will objectively show whether the accelerations are the same or not, I may ask whichever group that disagrees with the data why they think that the data is wrong. Then, if there’s enough time, I might suggest that they try a different setup that would circumvent the potential sources of error that they listed.

This LA focused on systematic error and asked the group to reflect explicitly on this to bolster their confidence in the result. This LA’s response is another example of allowing the experimental data to “speak” and allowing students to disagree with it only if they have a valid concern about the experimental design, data collection, or analysis. Potential student concerns about the data simply being “wrong” were addressed by giving students the opportunity to change their experimental setup and observe its effects on the data.

Establishing trust in the data was associated with a more closed start to the conversation, in which LAs did not elicit student ideas. This may be because LAs who allow the data to “speak” are concerned that, by opening up theoretical conversations about how the data fits within the physics schema developed through lecture early on, they may promote the sense that there is some “correct answer” to the lab, and sow doubt about any result that does not fit this answer.

Open-ended support of student conclusions. As noted, LAs taking pathway “A” tended to begin their responses with direct prompting of students that centered on evaluating data collection and analysis protocols, with little room for student extrapolation about what they “should” theoretically see in these results. However, the more narrow questioning approach LAs employed at the beginning of these pathway “A” encounters tended to give way to a more open and responsive approach than seen in answers from LAs taking pathway “B”, who tended to provide more targeted support to encourage students to come to a specific, supposed answer. This is to say, LAs taking pathway “B” may have led with a more open-ended approach, but ended with a more targeted prompting style as students neared a conclusion. We posit that this difference in the pattern of teaching moves may be associated with epistemological beliefs of LAs taking either pathway, since 60% of the pathway “B” LAs indicated a belief that there was one correct answer to the lab prompt while only 23% of the pathway “A” LAs indicated this belief in their response.

B. LA Barriers to Supporting Scientific Inquiry in Open-Ended Labs

In addressing our second research question, we used sample responses from LAs taking pathway “B” or “C” to unpack barriers that may have prevented LAs from taking pathway “A”. We found evidence of three barriers, but will address in this paper the most prevalent: LAs who carried into the lab with them a sense that there was one correct answer tended to take pathway “B”, guiding students to use the data to confirm an expected result. That is to say, LAs who believed that there
was one correct answer that students ought to reach often pursued verification-seeking procedures in their responses.

### Table 1. Breakdown of LA responses based on pathway taken and belief that the lab task has a correct answer. (n = 58)

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Total responses</th>
<th>Believe in correct answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Data-Driven</td>
<td>22</td>
<td>5</td>
</tr>
<tr>
<td>B: Theory-Confirming</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>C: Null</td>
<td>21</td>
<td>9</td>
</tr>
</tbody>
</table>

Across the 58 LA responses, 40% of responses (n = 23) showed evidence that the respondent believed there was one correct answer to the lab prompt. Table 1 shows the breakdown of LA responses by pathway and whether responses implied a belief in one correct answer. 60% of LAs who took a theory-confirming pathway (“B”) revealed that they believed there was one correct answer to the lab prompt, while only 23% of LAs who took the data-driven pathway (“A”) showed any sign of believing such a correct answer existed. A chi-square test performed between the data-driven and theory-confirming groups shows that LAs who took a data driven approach were less likely to believe in a correct answer to the lab task, $X^2(1, N = 37) = 5.268, \ p = .0217$.

LAs who took a neutral stance on the question of the “correct answer” tended to prime students to focus on their experimental design and data, and allow their data to inform their conclusion. An example of this approach appears in the following response, in which the LA said they would ask the disagreeing students to consider whether the answer to the lab prompt could depend on the way they initially chose to set up the experiment, such as the angle of the ramp:

*I would tell them to write out the free body diagram and to solve it first conceptually, then to use their measurements to support that claim. Once they drew the free body diagram, I would isolate the forces moving the iOLab in the +x -x direction. Then, after dividing by mass, we could look at just the accelerations. Then I would ask them to relate the accelerations going up and going down to each other. After that, once they realize that there is more acceleration on the way up, they can use this to relate to their measurements.*

Here, the LA centered physics to explain a desired result. It is not clear from the response what the LA planned to do if the student’s results fell outside of this expected outcome.

### IV. Conclusions

The aim of our study is to begin to explore the ways in which the practices of LAs in introductory labs may differ from LAs who teach in other contexts (e.g. discussion, lecture, office hours, etc.). As lab reforms in physics continue, we expect the implementation of LA programs in labs to increase as a means of enacting and sustaining these reforms. The scope of the work presented in this manuscript is limited, but we hope that it can be used to inform further studies of the practices of LAs and other instructors teaching in lab settings.

While we have identified one barrier to novice instructors’ implementation of lab pedagogy, we acknowledge that many other barriers may exist. Our analysis showed limited evidence of other barriers. For example, many novice LAs have not yet developed the pedagogical skills to determine when to “hold back” information that students can discover for themselves. In other responses, it is evident that many LAs lack confidence in their ability to perform error analysis, or may not philosophically understand its value. These two additional barriers will be further explored in future work.

In thinking about the preparation and training of LAs for lab instruction, it may be helpful to consider the work that has been done in effectively training graduate teaching assistants in lab instruction [27]. We recommend anyone who utilizes LAs as instructors in lab settings to ensure that their LA pedagogy course specifically confronts the topics of supporting student reasoning and error analysis in order to help LAs refine their understanding of these topics beyond the ideas they may have learned during their own experiences as students.

### Acknowledgments

We would like to thank Eric Kuo for feedback regarding the data analysis and Katie Ansell for directing us to relevant literature and providing feedback on the manuscript.


Applying Voting Theory to Mastery Grading; A Study of Faculty Interpretation of Course-Level Categorical-Score Distributions

M. T. Freeman,1 Amogh Sirnoorkar,2 James T. Laverty,2 and Bethany R. Wilcox1

1Department of Physics, University of Colorado, 390 UCB, Boulder, CO 80309
2Department of Physics, Kansas State University, Manhattan, Kansas 66506

The usefulness of a research-based assessment to an instructor can vary widely depending on how student performance on the assessment is presented. Currently, the Thermal and Statistical Physics Assessment (TaSPA) is being developed with a novel reporting method to offer targeted course-improvement strategies based on student performance rather than numerical student scores. This novel reporting method, however, brings with it unique challenges with respect to characterizing course-level performance. To address these challenges, we explore voting theory as a framework to assist us in understanding the implicit value judgements in how we decide on the feedback we generate for instructors. We have also surveyed faculty perception of course-level categorical performance distributions to learn about trends and areas of consensus in how faculty interpret performance distributions, which will inform what feedback TaSPA gives instructors based on their course performance.
I. INTRODUCTION

To improve students’ learning we must first be able to measure it. Within Physics Education Research (PER) this is primarily accomplished by research-based diagnostic tests. Considerable work is done to ensure these research-based assessments (RBAs) are reliable and valid with the goal that they can be used to improve courses by providing data on student learning directly to instructors. In practice, the onus of interpreting RBA data and deciding what to do with it is on instructors. However, making sense of the results of RBAs to inform concrete changes to classroom instruction can be challenging, even for the most practiced instructors [1]. This presents a barrier to the widespread use of these valuable tools as faculty have limited time and resources.

This work is situated within the development of a new RBA: the Thermal and Statistical Physics Assessment (TaSPA). TaSPA is currently being developed with specific attention to addressing the previously mentioned barrier by reporting results in a way that provides actionable feedback and includes suggestions for improving a faculty member’s course. Fundamentally, this feedback is based on a summary of the course-level performance (i.e. a single course performance metric). However, in order to be actionable, TaSPA does not just report a course mean and standard deviation, which can be difficult to interpret. Instead, TaSPA reports a course-level, categorical distribution of how many students achieved, partially achieved, or did not achieve the learning goal (see Sec. II). Ideally, this allows faculty to more clearly interpret their students’ performance [2]. This solution, while potentially allowing us to report results of RBAs in a more useful way, has its own unique challenges; when rating individual students into categorical bins (instead of a continuous score category), how do we generate feedback that meaningfully represents the course as a whole based on a distribution of categorical variables (i.e., how many students achieved, partially achieved, or did not achieve the learning goal)?

When interpreting these categorical distributions at a course level, a decision must be made about how to aggregate students’ individual performance categories into course-level metrics that describe the class performance as a whole. This decision significantly influences the degree to which these metrics are interpretable for instructors. Assessment research within PER has historically avoided this concern by reporting only students “scores” (i.e., a single continuous number meant to represent their performance on the assessment as a whole). For near-continuously distributed scores with a sufficiently large sample size, we can often describe the aggregate of these individual scores by a Gaussian distribution with a mean and standard deviation, then use that model to extrapolate characteristics of the full population of interest (not all of which is in our data set). However, in practice, single-course data sets are often not large enough to justify these assumptions, and, even when they are, these aggregate metrics can be hard to interpret (e.g., what do I change when my students got a 62% on average and what does that score mean?). Additionally, this approach assumes that score is a continuous variable. In the case of TaSPA the “score” a student receives is categorical; individual students’ “scores” are no longer on near-continuous scales.

This issue is not unique to TaSPA; it also arises for other fundamentally categorical schemes like Likert-style assessments as well as mastery grading [3, 4] more broadly (which also bases performance on a categorical scale of mastery or not). A problematic implication of this is that the aggregation of individual scores no longer has this clear parallel to Gaussian distributions in how to aggregate and interpret individual scores. This work seeks to address this challenge by drawing on another area – voting, which also deals with the challenge of taking individual categorical “scores” or votes and determining an appropriate overall choice meant to represent the will of the population (or, in our case, the performance of the course as a whole). In the following sections, we will lay out the specifics of TaSPA’s rating system, the background and specifics of voting theory, and operationalize voting theory for TaSPA with discussion of the specific challenges and questions that must be addressed in that process.

II. BACKGROUND & MOTIVATION

TaSPA was developed with specific attention to how faculty use assessment results to inform changes to their instruction based on theories of self-regulated learning and evidence-centered design (ECD) [2, 5–7]. TaSPA evaluates students with respect to their achievement of a learning goal categorically as: “Met”(M), “Partially Met”(P), or “Not Met”(N); referred to together as MPN categories. These categories are informed not only by ECD [8] but also by criteria developed for the Next Generation Science Standards [9], i.e. 3-dimensional learning [10]. To ensure this new format is useful to faculty, we previously solicited faculty feedback via interviews. This work suggested that faculty appreciate the format in which they can see the percentage of students in each performance category and the section of suggested course changes based on the overall course performance [11].

To implement this novel feedback system for TaSPA, we must decide how to turn individual categorical performance measures into a single course-level evaluation. To do so, it is worth explicitly stating our goals as they will steer us to some solutions over others (see Sec. III). TaSPA’s primary goal is to inform faculty and enable them to improve their teaching; thus, we posit that it is better to err on providing more feedback than less in cases where there is ambiguity. However, we also acknowledge that RBAs are sometimes used in an evaluative capacity (e.g., for tenure and promotion), and, thus, do not want to downplay strong performance.

As discussed in the previous section, the problem seen in interpreting TaSPA’s performance categories is also seen in Likert scales that are commonly used in research-based assessment instruments [12]. Historically, these instruments dealt with this challenge by aggregating individual scores in one of three ways: averaging responses on a linear scale (e.g.
In many cases the value judgements being made, and how they will affect aggregation and interpretation, are implicit, which is problematic when the effects of the judgements are not consistent over the aggregation population. To make these judgements more explicit, and to provide a set of analysis tools for these decisions, we can use a special case of Social Choice Theory (a set of frameworks for the aggregation of individual metrics into collective metrics) \cite{45} as applied to voting. This reframes our problem from one about a classroom of students to one of a population of voters. In this reframing, students become voters, and the metrics describing a class’s performance become the outcomes of this vote. This is a particularly useful perspective to take because a voting system can be used to produce class performance metrics that, by construction, embody the properties (the value judgements and effects of them) inherent to that system.

Much of voting theory is beyond the scope of this paper; however, the important aspect of it for this work specifically is the way that it deals with voting systems. In voting theory, a voting system (also called an “electoral system”) is a set of rules that describes how to take individual preferences and produce a group preference. These individual preferences may be a single choice or a ranked list. The group preference may be a single preference or multiple. This framework offers a lot of flexibility in how the preferences are aggregated and, as such, voting theory has some generalizable and powerful tools used for the analysis of these systems. One tool is the fairness criteria (an example of one is: if a preference has more than 50% support it should be selected as the group preference), which cannot all be satisfied for every possible vote by any electoral system (Arrow’s Impossibility Theorem) \cite{46}. This allows us to draw a parallel to implicit value judgements in how we aggregate individual student scores.

By casting our aggregation methods in the framework of voting theory these fairness criteria become the value judgments that allow us to weigh the possible violations of behaviours we might desire from our aggregation method.

\section*{III. VOTING THEORY}

For each class distribution given in the following questions, please select the response that most closely reflects the way you feel about the scale of changes you would make to how you teach that topic to improve future performance on it.

\begin{itemize}
  \item [Question 1 of 19]:
  \begin{itemize}
    \item M: 82% \\
    \item P: 13% \\
    \item N: 5% \\
    \begin{itemize}
      \item I don't need to make any changes to how I teach this topic in the course.
      \item I should make some moderate changes to how I teach this topic in the course.
      \item I need to make substantial changes to how I teach this topic in the course.
      \item None of the responses above are what I would say about this class.
    \end{itemize}
  \end{itemize}
\end{itemize}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig1.png}
\caption{The first question on the survey. Each question is on its own page with a short reminder of the instructions and has a different category distribution seen in the percentages below the question’s label. The first three response options will be referred to as “No Change”, “Moderate Change”, and “Significant Change”.}
\end{figure}

\section*{IV. METHODS}

This work’s research questions are: how do faculty interpret TaSPA’s category distributions, and what value judgments do faculty make while doing so. To answer these questions, we created and administered a survey to faculty that had taught or were teaching undergraduate courses, to analyze their responses with voting theory. The survey was designed to take 5 to 10 minutes to complete to minimize survey fatigue. The survey consisted of a preface, 19 short questions, and a demographics section. Each of the 19 questions ask the respondent to select how significantly they might change how they teach a topic in a hypothetical course based on a provided course-level category distribution. An example of this format can be seen in Fig. 1. The first 4 of the 19 questions were given to all respondents as controls, while the remaining 15 were pulled randomly from a larger set of 105 (see Sec. IV A). Faculty were asked to choose from options spanning ‘making no changes’ to ‘making substantial changes’ to their instruction, and a fourth option if they felt that their reaction to a distribution would not fit into the existing options.

The preface to the survey had the goal of putting the respondent into the mindset of someone who has just finished teaching a class where they administered a research-based diagnostic test that reported class performance broken down by the course topics. In this hypothetical situation, they have just received the results (in the format and language that TaSPA uses) for their class and must decide how significantly they might change how they teach these topics. The preface also explained the MPN categorization and explicitly addressed

This allows us to draw a parallel to implicit value judgements in how we aggregate individual student scores.

\section*{III. VOTING THEORY}

The sheer number of unique methods to aggregate Likert data is also an indicator that there are value judgements being made in the choice of aggregation method. An example of such a judgement for a Likert scale is that a linear scale scoring scheme values responses near the ends of the scale as discernible and uniformly removed from responses nearer to the center of the scale. This scheme places value on the extremity of a response whereas a scheme that bins the Likert scale in a binary fashion disregards the extremity and focuses only on the direction relative to the center of the scale.

In many cases the value judgements being made, and how they will affect aggregation and interpretation, are implicit, which is problematic when the effects of the judgements are not consistent over the aggregation population. To make these judgements more explicit, and to provide a set of analysis tools for these decisions, we can use a special case of Social Choice Theory (a set of frameworks for the aggregation of individual metrics into collective metrics) \cite{45} as applied to voting. This reframes our problem from one about a classroom of students to one of a population of voters. In this reframing, students become voters, and the metrics describing a class’s performance become the outcomes of this vote. This is a particularly useful perspective to take because a voting system can be used to produce class performance metrics that, by construction, embody the properties (the value judgements and effects of them) inherent to that system.

Much of voting theory is beyond the scope of this paper; however, the important aspect of it for this work specifically is the way that it deals with voting systems. In voting theory, a voting system (also called an “electoral system”) is a set of rules that describes how to take individual preferences and produce a group preference. These individual preferences may be a single choice or a ranked list. The group preference may be a single preference or multiple. This framework offers a lot of flexibility in how the preferences are aggregated and, as such, voting theory has some generalizable and powerful tools used for the analysis of these systems. One tool is the fairness criteria (an example of one is: if a preference has more than 50% support it should be selected as the group preference), which cannot all be satisfied for every possible vote by any electoral system (Arrow’s Impossibility Theorem) \cite{46}. This allows us to draw a parallel to implicit value judgements in how we aggregate individual student scores.

By casting our aggregation methods in the framework of voting theory these fairness criteria become the value judgments that allow us to weigh the possible violations of behaviours we might desire from our aggregation method.

\section*{IV. METHODS}

This work’s research questions are: how do faculty interpret TaSPA’s category distributions, and what value judgments do faculty make while doing so. To answer these questions, we created and administered a survey to faculty that had taught or were teaching undergraduate courses, to analyze their responses with voting theory. The survey was designed to take 5 to 10 minutes to complete to minimize survey fatigue. The survey consisted of a preface, 19 short questions, and a demographics section. Each of the 19 questions ask the respondent to select how significantly they might change how they teach a topic in a hypothetical course based on a provided course-level category distribution. An example of this format can be seen in Fig. 1. The first 4 of the 19 questions were given to all respondents as controls, while the remaining 15 were pulled randomly from a larger set of 105 (see Sec. IV A). Faculty were asked to choose from options spanning ‘making no changes’ to ‘making substantial changes’ to their instruction, and a fourth option if they felt that their reaction to a distribution would not fit into the existing options.

The preface to the survey had the goal of putting the respondent into the mindset of someone who has just finished teaching a class where they administered a research-based diagnostic test that reported class performance broken down by the course topics. In this hypothetical situation, they have just received the results (in the format and language that TaSPA uses) for their class and must decide how significantly they might change how they teach these topics. The preface also explained the MPN categorization and explicitly addressed
concerns or issues faculty might have with the survey (identified in a pilot administration of the survey). To do this we: explicitly stated the goal of the survey as “to investigate faculty interpretation of our novel reporting format”; acknowledged that teaching is complex, but noted we are asking them to imagine the distributions are the only information available to base their changes on; and acknowledged that some distributions would seem similar but were designed to probe for edge cases and nuances in interpretation.

A. Question Design

The first 4 questions of the total 19 given were the same for all respondents. This had two main purposes: to see how respondents agreed or disagreed on the same distributions, and to see how respondents interpreted different extreme distributions (e.g., distributions that are close to being entirely in one of the MPN categories). Three of these control questions (C-N, C-P, C-M in Fig. 2) were randomly generated to be in the extreme cases of high M or P or N. The last control (C-Cent.) was picked to be roughly evenly distributed in the categories as we were interested in how faculty would interpret this particular point in the distribution space.

The remaining 15 questions were pulled from 15 groups of 7 questions. Each of the 7 points (questions) in each group was centered on a randomly generated MPN distribution with a spread of 6 more points around it (see Fig. 2). This was done by taking the MPN percentages and pushing the distribution “towards” (adding 4% to one category and subtracting 2 from the others) or “away” (subtracting 4 from one category and adding 2 to the others respectively) from the extremes. The size of the spread around the central distribution was chosen to keep the percentage difference low but still meaningful. This clustering around a central distribution allowed us to get an idea of how respondents would have responded differently to small local changes in our choice of distributions as well as a sense of agreement in a region around a central distribution.

From each of the 15 groups of 7, each respondent saw one randomly selected distribution such that the 7 were equally given to respondents. The 15 central distributions were chosen as a compromise between wanting more coverage over possible distributions and survey length. These competing desires led us to 15 randomly generated center distributions by starting with approximately 20 and then removing questions that were in the most densely populated regions of the distribution space. This left us with 15 random questions spread over the possible distributions with 6 additional questions around each. These 105 total distributions were the final ones we settled on in addition to the 4 controls (see Fig. 2).

The demographics questions at the end of the survey asked for: the type of institution respondents belonged to, how many years they had been teaching at the undergraduate level, how many years since they last taught at the undergraduate level, what their typical class size was, and how often they used RBAs. These questions were chosen to provide additional insight into the background of individual respondents.

B. Survey Context

The survey was administered in the Spring of 2023 and gathered 33 responses, 9 of which were incomplete (the faculty member did not answer all questions), and 1 that was completed but all 19 questions had the 4th response option selected (“None of the responses above are what I would say about this class”). These 9 responses are not included in analysis. Of the remaining 23 faculty, only 3 ever selected the 4th response option. The maximum number of times an individual respondent selected the 4th option was 6 times, and on no question did these 3 faculty unanimously select the 4th option. Thus, the 4th option is excluded from our analysis as it did not occur often enough to support clear trends. All faculty self-identified as belonging to a “4-year Research Focused” institution and had a minimum of 6 years teaching experience. Of all faculty, 87% reported that they were currently teaching undergraduate courses, all others had taught an undergraduate course at most 2 years prior. Class size was highly variable, with 60% of respondents teaching classes of 10 to 110 students and the next largest population being faculty teaching classes typically larger than 200 students. In response to being asked to select how often they have used RBAs, faculty had the option to select “Always” (N=8), “Sometimes” (N=6), “I’ve tried them but no longer use them” (N=6), “Never” (N=1), and “I was not aware of such tests prior to this survey” (N=1). One respondent selected “Other” writing “Always if available and can be scored meaningfully.”

V. RESULTS & DISCUSSION

Responses to the survey are plotted in Fig. 2 (color mixing described in the caption). “No Change” was selected most for C-M (74%); Q18 had the next largest percentage of “No Change” (30%). Without additional questions in the region between C-M and Q18, we cannot infer whether the change in this percentage drops off suddenly or smoothly, but it is notable that even a question with a 80% Met did not get unanimous consensus from the respondents that no changes to instruction would be necessary. To identify consensus (or not), we took every group of seven points and collapsed the faculty response for each group into a single “point” by identifying the option selected by the largest number of respondents for each distribution in that group. The collapsed “points” were then evaluated based on how many faculty agreed with this collapsed evaluation. This “agreement” is equivalent to how much consensus there is between faculty on their interpretation of distributions within each grouping. All groupings except 6 achieved an agreement of 67% (i.e., a supermajority) or more indicating general consensus amongst our respondents.

Part of the goal of this work is to inform the development of feedback for TaSPA to ensure it is consistent with faculty interpretations of our score distributions. To establish a qualitative model of faculty consensus, we have grouped the responses into 5 regions denoted by the gray dashed and solid lines seen on Fig. 2. These lines were drawn qualitatively
so that they did not go through any points and separated the low-agreement groups (the 6 mentioned previously) from the high-agreement groups. Bounded by solid lines is the region that should contain the crossover between Moderate and Significant Change, while the dashed lines should contain the crossover between Moderate and No Change. This splits space into 5 regions termed 1 through 5 in Fig. 2; in these regions we have: agreement on Significant Change; ambiguity between Moderate and Significant Change; agreement on Moderate Change; ambiguity between No Change and Moderate Change; and agreement on No Change, respectively. Note that the large gap in points between regions 4 and 5 make the positioning of the dashed line dividing them highly uncertain, though erring on the side of more changes (or giving more feedback, see Sec. II) means putting the boundary closer to C-M. This point is discussed further in Sec. VI.

To evaluate the degree to which this simple region model fits the data, each faculty response was compared to the model and the model was scored based on number of points: correctly predicted (i.e. Significant, Moderate, and No Change in regions 1, 3, and 5 receptively); incorrectly predicted (i.e. No Change in regions 1, 2, or 3; Moderate Change in 1 or 5; or Significant Change in 3, 4, or 5); and unpredicted (i.e. any point in regions 2 or 4 that is not incorrect). The percent of points in each of these categories was calculated per respondent and then averaged. The average correct rate was 59%, the average incorrect was 16%, and the average unpredicted was 25%. The correct and incorrect percentages indicate how often the model was “right”; however, the interpretation of the unpredicted category is nuanced. It is tempting to assume we want this number to go to 0%, but that would necessarily increase the incorrect category due to the inherent variation in faculty responses. Thus, the goal of this category is to account for the variance in perception between faculty of these distributions in our model such that we can decrease the number of incorrect responses without over-fitting our data.

VI. CONCLUSIONS & FUTURE WORK

Here, we explored faculty interpretations of a novel approach to reporting student performance on research-based assessments. Based on their responses, we created a simple region model to convert categorical distributions of student performance to course-scale feedback that can help faculty decide whether to make changes to their instruction. As a preliminary study with small N, this simple region model provides a useful tool for understanding faculty interpretations to inform development of useful feedback. In particular, this work will guide us in developing the algorithm that generates feedback reports for TaSPA based on class MPN distribution, ensuring feedback aligns with faculty interpretation. This work also allows for investigation of how faculty might interpret these categorical distributions and allows us, as assessment developers, to gain a better understanding of our audience and where our priorities and values might meaningfully differ. For example, the general trends in our data support the claim made earlier that, in general, faculty will prefer to make changes over not making changes, even for classes with performance levels an assessment developer might consider quite high (e.g., >70%). This supports the goal that TaSPA sets of erring on the side of providing more feedback to instructors.

While this work was motivated as part of the broader TaSPA project, because this survey was not specific to TaSPA, this procedure could be used to investigate aggregation of other categorical scales, such as mastery grading and Likert-style. Future work for this project could include developing a modified survey with more questions centered around the regions of ambiguity identified in this round and targeting a larger faculty population. This increased statistical power would allow us to identify existing voting systems that predict faculty responses. Identifying an appropriate voting theory would provide a way to look for inherent value judgments faculty, as a collective, are making when interpreting categorical scoring such as mastery grading or Likert-style questions.

ACKNOWLEDGMENTS

Thank you to Sara Moore, who introduced me to voting theory, and the TaSPA team who provided feedback on this project. This work was supported by the National Science Foundation under Grant No.s 2013332 & 2013339.


[18] Edward F. Redish, Jeffrey Saul, and Richard Steinberg, “Stu-


[21] Richard L. Pearson III, Savannah L. Logan, and Wendy Adams, “Faculty perception insights obtained from faculty inter-


[25] Karen Cummings, Stephanie Lockwood, and Jeff Marx, “Atti-


[27] Karen Cummings, Stephanie Lockwood, and Jeffrey Marx, “Attitudes toward problem solving as predictors of student suc-


[31] Christine Lindstrøm and Manjula Sharma, “Physics self-


[40] Christine Lindstrøm and Manjula Sharma, “Physics self-


[49] Edward F. Redish, Jeffrey Saul, and Richard Steinberg, “Stu-


[52] Richard L. Pearson III, Savannah L. Logan, and Wendy Adams, “Faculty perception insights obtained from faculty inter-


[56] Karen Cummings, Stephanie Lockwood, and Jeff Marx, “Atti-


[58] Karen Cummings, Stephanie Lockwood, and Jeffrey Marx, “Attitudes toward problem solving as predictors of student suc-


[60] Ben Zwickl and Heather Lewandowski, “Colorado learning at-


Broadening Student Learning through Informal Physics Programs

Carlee Garrett and Tatiana L. Erukhimova
Department of Physics & Astronomy, Texas A&M University, College Station, Texas, 77843

Jonathan D. Perry
Department of Physics, University of Texas at Austin, Austin, Texas 78712

Jonan P. Donaldson
Center for Teaching Excellent, Texas A&M University, College Station, Texas, 77843

Student learning at the university level occurs in a variety of settings, both formal and informal. Prior research shows that retention of knowledge is enhanced when students teach material related to their learning. In this study, we explored student perceptions of learning experienced through facilitation of informal physics programs, also called outreach, where they frequently practice methods of self-explanation to diverse audiences. To characterize the impacts of these facilitation experiences, we employed a student-centered investigation drawing on self-reported data gathered through didactic interviews conducted with 35 students who facilitated at least one of five informal physics programs. Analysis of interviews drew on multiple learning theories to characterize perceptions of understanding of physics concepts, confidence in their knowledge, and how those constructs related to engagement with members of the public through outreach. Using network analysis, we found three distinct clusters of themes focusing on disciplinary learning, internal development, and external engagement.
I. INTRODUCTION

Informal physics programs, also called physics outreach or public engagement programs, have the potential to enhance the educational experiences of university students. Recent research shows that students who were involved in facilitation of informal physics programs reported a positive development of their disciplinary identity and sense of belonging to the physics community [1, 2]. These are important factors that could help retention efforts in physics, especially for underrepresented groups [3–5]. Other important student experiences with informal physics programs include teamwork opportunities and improvement of student communication, presentation, and design skills [2, 6]. This development of important career skills may be achieved through student explanation of the physics concepts to the general public or team members and fabrication of new physics demonstrations [2, 7].

Students who facilitate informal physics programs are frequently engaged in methods of self-explanation to diverse audiences [2, 8]. Self-explanation is a method of actively processing new information by restating it in more familiar terms, relating it to existing knowledge, and making inferences [9, 10]. Prior literature indicates that students learning of new concepts deepens when they are expected to explain the concepts to other people [10]. Moreover, Fiorella et al. states that “learning is enhanced through the act of teaching others”; that is students tend to develop a deeper and more persistent understanding of material when they present a lesson, rather than just preparing to give a lesson [11].

Students who have the opportunity to teach have traditionally done so through positions such as being a teaching assistant or learning assistant. Informal physics programs provide university students with rich opportunities for teaching and learning in environments that are less structured, lower stakes, and often more exciting than regular classes. By design, these programs have more flexibility for student ownership, creativity, and innovation and have to be studied more thoroughly since many universities and colleges in the United States are running informal physics outreach programs. A study of one after school physics program facilitated by University of Colorado, Boulder students provided the community with findings demonstrating a positive influence of this experience on university student perspectives on teaching and learning [7]. Subsequent studies of the same program identified and refined an instructor pedagogy in informal physics learning environments [8, 12].

This study focused on exploring the impacts on student perceptions of their learning that facilitation of informal physics programs provides. We predicted that students who teach physics to public audiences through informal programs would report enhanced learning of concepts and synthesis of physics knowledge. In the following sections, we present our methods, including the theoretical framework, results and discussion, and end with some concluding remarks.

II. METHODS

To explore the learning that may occur through the facilitation of informal physics programs, we analyzed a series of 35 interviews, 11 women and 24 men, collected as part of a prior mixed-methods study [2]. Interviews were conducted with undergraduate and graduate students who facilitated at least one informal physics outreach program between 2013-2019 at Texas A&M University, a large, land-grant institution. Facilitators were engaged with at least one of five informal physics programs ranging from large annual festivals to smaller engagements with the general public happening throughout the year. Each interview was semi-structured, conducted by a researcher unfamiliar with the interviewee, and consisted of a set of six didactic questions developed in collaboration with learning scientists. These questions probed interviewees on how they saw their experiences with facilitating informal physics programs relating to constructs such as physics identity, worldview, and goals [5].

For this study, we related physics learning to growth in disciplinary knowledge as well as a broader process of change within students. These changes were expressed in multiple ways, including: transitioning from novice to expert perspectives, identifying as a member of the physics community, changing assumptions about the responsibilities of physicists, transforming due to authentic experiences, and encouraging physics identity development. To account for the complexity of learning experiences, we developed a code book based on four fundamental learning theories. The theories we drew on included social constructivist theory, situated learning theory, transformative learning theory, and constructionist theory.

Social constructivist theory asserts that knowledge is constructed through an active, social process where learners make connections between their background and new information [13, 14]. Codes drawn from this theory related to dialogue between facilitators and audience, focused on negotiating meaning, scaffolding, and identifying the zone of proximal development. Situated learning theory connects learning to interactions, relationships, engaging in the practices of a community, and an overall sense of transitioning from a novice to an expert through legitimate peripheral participation as newcomers grow in to experts [15, 16]. Related to this theory, codes for peripheral participation, as well as physics identity, ways of knowing, and disciplinary practices as they related to self, peers, and the audience were included. In defining physics identity, we drew on the physics identity framework developed by Hazari and colleagues, which includes interest/motivation, performance and competency beliefs, and internal/external recognition [3]. Transformational learning theory is concerned with changes in a learner’s “frame of reference”, or perspective, worldview, or assumptions [17, 18]. Codes related to the questioning and changing of assumptions, beliefs, or perspectives connected to physics were included from this framework. Constructionist theory frames learning as being tied to the creation of artifacts, activities, processes, or other observable things which
were reflective of the material related to what is being constructed \([19, 20]\). Codes related to this theory included making, designing, or building, as well as authentic audience, purpose, or impact which relates to activities and processes. Other learning theories were considered but omitted from this study as they did not describe the types of learning observed through informal physics programs. We also included important outcomes related to learning, such as career skills and future plans. A total of 23 codes were used in this study.

After developing the code book, three researchers split in to two teams and coded each interview while meeting regularly throughout the process to discuss and resolve differences. For this work we treated complete sentences as our fundamental unit, the smallest portion of interviews which could contain distinct ideas representing our codes. After completing this process, the two teams had an intercoder agreement of \( \kappa > 0.9 \). We examined the relationships between codes through network analysis. Network analysis provides a visual representation of statistically significant relationships between nodes (in this case codes), as well as highly interrelated clusters of nodes. By calculating correlation matrices and determining the centrality of each idea through eigenvector measures, a map is produced where edges (lines) show statistically significant connections at an assigned level, \( p < 0.01 \) for this study. Larger nodes and a higher number of edges are related to the eigenvector centrality of a node within the framework \([21, 22]\). Clusters within the map are designated by different colors and shapes of nodes, where clustering was done with a Girvan-Newman cluster analysis \([23]\). The clusters produced by this analysis are considered robust and an accurate representation of relationships when a value, termed \( Q \), is above 0.30. Each cluster represents a group of interrelated nodes which are more strongly linked together than they are linked to nodes in other clusters. In other words, clusters represent a subset of interdependent ideas.

### III. RESULTS & DISCUSSION

In this section, we present initial results from our analysis of interviews with undergraduate and graduate students who facilitated one or more informal physics programs, and discuss their implications. Results of a network analysis, at the \( p < 0.01 \) level, shown in Fig. 1, include 17 out of 23 codes which are grouped into three distinct Girvan-Newman clusters \((Q = 0.423)\). The remaining codes were not correlated with others at the level of statistical significance used for this study. Three clusters were identified. Based on our interpretation we labeled them as: disciplinary development, internal development, and external engagement. Each of these clusters is discussed below.

The first cluster we will discuss is represented by the black triangles in Fig. 1. Within this cluster we observe themes around disciplinary development. Through working within informal physics programs, students reported development of their physics knowledge and ways of knowing. Facilitators attributed informal physics programs as an environment where they developed their disciplinary knowledge beyond the classroom. Students shared that "getting the hands-on experience" of working with demonstrations and "knowing it well enough that you can...teach other people with it was instrumental to me in realizing that I knew physics." These findings mirror prior research which has shown that students benefit from being in a teaching position as they develop and refine an explanation over a topic, leading to deeper learning and increased retention of the content they taught \([24]\).

Within our network, these developments were related to students’ current degree plans and opportunities to create new demonstrations and share them with audiences. When working in a program to design or build new demonstrations, students were placed on a team with other members of varying levels of knowledge, from undergraduate freshmen to graduate students. One student shared that his team developed a bond as they would discuss “anything from...how they became physicists, to what their study is,” exposing each other to their individual fields of interest. That student also shared that they were offered a “research opportunity [that] would not have come about if [they were] not a part of the [program] and doing outreach”. Through informal physics programs, students formed meaningful connections with others in their discipline. From having unstructured interactions with others, students were also able to expand their physics knowledge, and refine their current degree plans.

The second cluster we will discuss is represented by the blue squares in Fig. 1. These codes are more focused on the internal development of facilitators. These interdependencies highlight how students have evolved through their facilitation of informal physics programs individually and through peer interactions. In multiple interviews, students reported that they have developed their sense of responsibility through their engagement with informal physics programs. As one student shared, working with a team as part of an informal physics program was when they had a sense others looked at them “as someone that should know what they were doing.” Additionally, facilitators reported their beliefs regarding academic responsibilities evolved to include a more comprehensive view of their opportunities. One student mentioned that getting to interact with peers and faculty during informal physics programs led to them fully realizing that there is not one set academic path which works for everyone pursuing a physics degree. Instead, “there are other routes, and they’re not any better or worse” than one another. Another student commented that after hearing their peers discuss their ideas and goals, they felt better exposed to opportunities available to them, stating that they "didn’t know that [those goals] were things you could do". These experiences were reported to have influenced the participant’s perspective of themselves as they progressed from being seen as a novice practitioner to holding a position of greater expertise. This was related to participants’ learning of nondisciplinary qualities such as leadership and communication skills. Students also reported on refined ideas of what they would pursue after comple-
FIG. 1. Network map of relationships between learning and outcome codes at the $p < 0.01$ level. The size of the nodes indicate their Eigenvector centrality. Colors and shapes of the nodes denote distinct clusters, $Q = 0.423$, of statistically interconnected ideas representing more closely related themes as determined by a Girvan-Newman clustering analysis.

ing their current degrees. One student in particular shared how they knew they wanted to continue working with children because of “this moment of clarity” where they better understood “who [they] wanted to be” professionally. In relation to physics identity, facilitators reported direct impacts from their experiences in outreach to improvements in their interest and motivation to learn more physics, as well as their confidence and competency beliefs in what they had already learned. From one student’s experience, they noted that these changes were “mostly due to me having that confidence presenting outreach demos and just having that little surge of confidence every week to go into my academics and presentations.” Overall, interviewees shared that informal physics program opportunities provided an environment for them to engage with others in and outside of their discipline, which created the necessary authentic setting needed to transform their perceptions of their physics identity.

The third cluster in Fig. 1, indicated by the red circles, represents factors related to external engagement. These factors revolved around facilitators discussing their development of skills in engaging with audiences during informal physics programs, as well as evaluating the effectiveness of their explanations. One student shared how “physics outreach really forces you to be able to explain an idea in a variety of levels” since “you end up talking to people with multitudes of backgrounds and knowledge.” Through this student’s experience being a facilitator, they have seen “people who have no understanding” and “some people show up, and then they’re like, ‘Well, actually-’ and you’re like, ‘Yeah. I know. That’s right.’”. This encouraged negotiations between the facilitator and a diverse audience to establish the audience’s current knowledge and allow the facilitator to scaffold their content and effectively communicate. It should be noted that the questioning process and direct teaching motivates students placed in teaching roles to process the material at a deeper level. Prior research has discussed asking and answering questions as an advantage when it comes to students learning because it requires them to look at the material from multiple perspectives [10]. These conversations between student facilitators and the public also encouraged audiences to participate in the demonstrations and develop science-like practices and ways of thinking. Facilitators identified these interactions as developing audience members physics identities while they progressed towards their zone of proximal development and reached a deeper understanding of what was being presented. One student commented on a time where “a little girl… came up to [her] after explaining one of the big explosion demos and she was just like, ‘You are so cool. How do you explain that? Where did you learn all this stuff?’” The way that this facilitator interacted with the audience encouraged interactions, as shown by the audience member asking multiple questions. Lastly, facilitators noted that developing their physics identity also impacted the audience’s ways of knowing. As exemplified by one facilitator, they have seen the audience “want to see how lasers work… they want to see hover crafts and all that.” And when the student shared their expla-
nation, the audience “intuitively get the concepts behind it.” Thus, the facilitators asserted their belief that when the audience appeared more motivated and engaged with the physics behind the demonstrations, they appeared to process the information that facilitators shared more effectively.

While each cluster represents an important theme of learning, there are also connections between each cluster which represent important links between the themes discussed above. Based on our analysis, connections between disciplinary development and internal development formed as students progressed in their current degrees and their peripheral participation. Informal physics experiences allowed students to feel more a part of the field, growing from novices to experts, and for some resulted in higher certainty and more refined plans for completing their degrees, such as engaging in research efforts as mentioned previously.

Disciplinary development and external engagement formed connections from facilitators reflecting on their ways of knowing while scaffolding information for the audience members. By interacting with "people who are not 'physics people" inherently, facilitators were able to check their own understanding along the way as they built upon the audience’s understanding by scaffolding the needed information. Informal physics programs provided ample opportunities for students to scaffold their knowledge for audience members ranging from 5 to 95 years old and from those new to physics to those with significant knowledge. Through these dynamic interactions, students reported improvements to their understanding of topics covered in classes beyond the level of simply being able to pass exams. Being a facilitator "helps [the students] interact with people who might be able to correct things that [the facilitator] didn't understand," making the interaction mutually beneficial.

Internal development and external engagement connect along three different edges. The first occurred with facilitators recognizing their peers physics identity development while also evaluating the audience’s ways of knowing. Students reported that when observing their peers presenting demonstrations they simultaneously noted the audience developing their ways of knowing in receiving the information, and their peers growing in their physics identity from discussing physics. The student’s peers got to demonstrate fundamental components of physics and "connect [them] to the broader research...like in [their] own labs” for the audience. This dialog with a diverse audience, which at times led to achieving a common understanding, was seen by researchers to be related to growth in students’ physics identity. Internal development also crossed into external engagement when facilitators reported on their own physics identity development while either negotiating meaning with the audience or encouraging the audience to engage in physics practices via hands-on demonstrations. Students reported that delivering their explanations was "a really good form of practice" and they got "exposure to the way people think” through discussing the physics concepts behind the demonstrations with audience members. Additionally, these negotiations "translated to the classroom" and contributed to the students’ development as physicists, engaging them in their physics identity development. Leading the audience through understanding a demonstration helped students feel most like a physicist and enhanced their feeling of expertise.

IV. CONCLUSIONS

This study expands on prior analyses of student experiences gained through facilitation of informal physics programs to analyze perceptions of student learning, an important dimension of their experience as undergraduate or graduate students in physics. Our examination exhibited three distinct clusters or types of learning: disciplinary development, internal development, and external engagement. By engaging with programs that brought students into contact with diverse audiences in less structured settings, they were able to broaden and deepen their understanding of physics content both through the act of creating new demonstrations, and engaging in dialogue with those more novice than themselves to explain physics concepts. Through feedback from audiences, their peers, and self-reflection, facilitators also built an awareness of their transition from more novice to more expert physicists. Statistically significant links between these three clusters indicate that the types of learning and growth reported in the previous section are intertwined. In other words, they are all features of experiences gained through informal physics programs and may not be separable.

While these findings are interesting and contribute to the growing body of literature about the benefits of university students facilitating informal physics programs, we must note some limitations in this study. All self-reported data were collected from a single institution. Students who agreed to be interviewed were self-selected volunteers. Future work in this area will build on these findings and draw from data which includes a larger body of responses from students who did and did not engage in facilitation of informal physics programs.

ACKNOWLEDGMENTS

We would like to thank Callie Rethman and Daniel Choi who conducted the interviews as part of the original study. We would also like to thank all students are gave their time to sit for the interviews. This work was support in part by the Texas A&M University College of Arts and Sciences.


Representational differences in how students compare measurements

Gayle Geschwind\textsuperscript{1,2}, Michael Vignal\textsuperscript{1,2}, and H. J. Lewandowski\textsuperscript{1,2}

\textsuperscript{1}Department of Physics, University of Colorado, Boulder, Colorado 80309, USA and
\textsuperscript{2}JILA, National Institute of Standards and Technology and University of Colorado, Boulder, Colorado 80309, USA

Measurement uncertainty plays a critical role in the process of experimental physics. It is useful to be able to assess student proficiency around the topic to iteratively improve instruction and student learning. For the topic of measurement uncertainty, we developed an assessment tool called the Survey of Physics Reasoning on Uncertainty Concepts in Experiments (SPRUCE), which aims to assess students’ knowledge, and use of, a variety of concepts related to measurement uncertainty. This assessment includes two isomorphic questions focused on comparing two measurements with uncertainty. One is presented numerically and the other pictorially. Despite the questions probing identical concepts, students answer them in different ways, indicating that they rely on distinct modes of representation to make sense of measurement uncertainty and comparisons. Specifically, students score much higher on the pictorially represented item, which suggests possible instructional changes to leverage students' use of representations while working with concepts of measurement uncertainty.
I. INTRODUCTION & BACKGROUND

All measured quantities have associated uncertainties, making measurement uncertainty a crucial aspect of experimental physics. Using measurement uncertainty correctly is essential for interpreting measurements, presenting results, and drawing reliable conclusions based on those results. The Effective Practices for Physics Programs (EP3) Guide [1] also emphasizes the significance of learning measurement uncertainty techniques as taught in physics laboratories. Despite its critical role, students frequently struggle with concepts and practices surrounding measurement uncertainty, including propagation of error, comparison of measurements, calculating standard deviations and standard errors, and taking several measurements to get a distribution of results, even after taking a course emphasizing these areas [2–7].

As part of efforts to improve student learning of measurement uncertainty, we have developed a new research-based assessment instrument (RBAI) called the Survey of Physics Reasoning on Uncertainty Concepts in Experiments (SPRUCE) [8, 9]. SPRUCE is an online assessment intended to be utilized in a pre-post format allowing instructors to measure the impact of a course on students’ proficiency with concepts and practices of measurement uncertainty. We developed SPRUCE using the framework of Evidence-Centered Design (ECD) [10], a robust method of creating and validating an RBAI. Although validation is an ongoing project (with a future paper in progress), SPRUCE still offers a wide variety of insights into how students handle measurement uncertainty. Its design provides instructors with their students’ progress along 14 dimensions referred to as Assessment Objectives (AOs) [11] after one term of a laboratory class.

AOs are “concise, specific articulations of measurable desired student performances regarding concepts and/or practices targeted by the assessment [11].” AOs are similar to course learning goals and are essentially the constructs the assessment aims to measure. We developed the SPRUCE AOs with input from introductory laboratory instructors to determine which aspects of measurement uncertainty they find important and want their students to learn in their courses [8]. These AOs then aided in writing the SPRUCE assessment items; each item on SPRUCE addresses at least one of these objectives. In this way, we focused the scope of creating SPRUCE to topics instructors frequently deem important to their introductory laboratory courses.

Here, we examine one particular SPRUCE AO: Determine if two measurements (with uncertainty) agree with each other. SPRUCE has two isomorphic questions for this objective. First, the assessment presents students with numerical measurements and asks about agreement between these measurements. Then, later in the assessment, with several questions in between, a similar question appears with the data represented pictorially, as symbols with error bars. Students are not explicitly informed about the relationship between these two items. This allows us to probe how students are able to compare measurements when presented with the same data with two different representations.

<table>
<thead>
<tr>
<th>Number of Institutions</th>
<th>Institution Type</th>
<th>Number of post responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 Year</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>4 Year</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>Master’s</td>
<td>39</td>
</tr>
<tr>
<td>5</td>
<td>PhD</td>
<td>617</td>
</tr>
</tbody>
</table>

Existing literature has explored the use of multiple representations while students problem solve [12–16]. For example, Kohl et al. found that students frequently view a mathematical problem and a pictorial problem as ‘opposites,’ where students consider pictorial problems as more aligned with “concepts,” which are frequently treated distinctly from numerical problems. Further, they found statistically significant differences in performance based on different representations of isomorphic problems on homework and quizzes. Students tended to perform worse on problems in a mathematical or numerical format than with problems in other formats (e.g., pictorial, verbal, or graphical)[12].

The work presented here aims to identify whether student performance in comparing measurements similarly depends on representation. To do this, we will answer the following research questions.

- Do students respond differently to questions about comparing measurements when presented with different representations?
- How do students reason about comparing measurements when presented with different representations?

II. METHODOLOGY

We use a mixed methods approach, as the data collection and the analysis involve qualitative and quantitative components. To study students’ handling of measurement uncertainty, we administered SPRUCE in a pre-post online format during the Fall 2022 semester in 12 courses at eight institutions (See Table I). We received 670 valid post-instruction responses after we removed responses from students who did not consent to have their data used for research, did not correctly answer the filter question, or did not answer both items of interest.

We also conducted interviews during the Fall 2022 semester. These interviews aimed to determine whether students interpreted all of the items on SPRUCE as intended, as well as to probe student reasoning for each answer option on the assessment. Students were recruited from seven courses at four institutions (2-year, Master’s and PhD granting) already participating in the administration of SPRUCE during this semester. Each of the 27 interviews conducted lasted approximately one hour and students were compensated for their time. Interviewers (two of the authors) observed as students completed SPRUCE and inquired about students’ reasoning for each answer selected, as well as about why they did not select certain answer options. The interviews were audio/video recorded for future reference. Analysis of these
interviews consisted of taking notes during interviews and transcribing student quotes as needed.

For the analysis, we focused on the responses to two isomorphic multiple-response items, focusing on both the difficulty [17, 18] of these items and student reasoning for their answers to both. The first item, as shown in the upper half of Fig. 1, presents students with their ‘own’ numerical data (with uncertainty) for a measurement of a spring constant; they are then asked to select all answer choices of numerical data (means with uncertainties) that agree with their measurement. The second item presents these similar data in pictorial form, as shown in the lower half of Fig. 1. For brevity, we will refer to the numerically represented item as NRI and the pictorially represented item as PRI for the remainder of this paper. Students receive credit on these multiple response items by answering with the combination ‘ABCD’ or ‘ABCDF,’ based on expert responses. The uncertainties in both items represent the standard error; therefore, overlap or near overlap of the error bars is required for agreement. No other answer combinations earn credit, and no partial credit is awarded for these items. Note that we changed the order of answer options for the PRI for this paper to make discussion of the items easier.

III. RESULTS & DISCUSSION
A. Overall difficulty scores

While laboratory instruction commonly focuses on measurement comparison [8], low scores on both of these items at the end of the term indicate persistent student difficulties in handling comparison with uncertainties. Students score an average of $(25 \pm 3\%)$ on the NRI and an average of $(40 \pm 4\%)$ on the PRI on the post-test, with the error indicating 95% confidence interval. These scores indicate that, while not many students answered these items correctly, students answered the PRI correctly more often. We conducted a Mann-Whitney U test (a nonparametric test for independent measures) [19] to determine if this represents significant statistical difference, and found the p-value for these items as $p = 2.1 \times 10^{-8}$, indicating a statistically significant difference in student performance on these items. Additionally, we calculated the effect size to compare these two items using Cohen’s $d$ [20, 21], finding $d = 0.31 \pm 0.05$, showing a moderate effect size.

We also calculated the Pearson coefficient to determine the correlation between the two items. The Pearson coefficient varies between $r = -1$ and $r = 1$, where a more positive coefficient indicates a stronger positive correlation [17]. Anything above $r \approx 0.30$ indicates a fairly significant positive correlation. For these items, we find $r = 0.45 \pm 0.04$, which shows a fairly significant correlation in that if a student correctly answered one item, they are more likely to have correctly answered the other. However, the correlation is not perfect ($r = 1$): many students correctly answer only one of these items. The number of students who answered each question correctly is presented in Table II. Only about half of the students who correctly answered the NRI also correctly answered the PRI. This suggests that students who are able to reason through the numerically presented data seem better equipped to handle the pictorially presented data, but the reverse is not true on average.

We turn to the qualitative interview data to help to un-
TABLE II. Number of students who answered the NRI, PRI, or both correctly [N = 670]; error shown as 95% confidence interval

<table>
<thead>
<tr>
<th>Only NRI</th>
<th>Only PRI</th>
<th>Both Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Students</td>
<td>38</td>
<td>134</td>
</tr>
<tr>
<td>Percent of Students</td>
<td>6 ± 1</td>
<td>20 ± 3</td>
</tr>
</tbody>
</table>

understand these results. During interviews, some students describe mentally switching from a numeric to a pictorial representation easily and using this skill to solve the numeric item:

> I just looked at the values and saw it – like I kind of picture if they have that little bar with their error bars to see if they overlap.

This student essentially converted the numeral data into pictorial data in their mind and then used that representation to reason about the comparisons. Using this skill of mentally changing representations, they were able to answer both items correctly. This finding is similar to ones from Kohl et al. [12] and Welieferiya et al. [22], in which students were often able to switch between different representations when forming mental models of data.

### B. Individual answer analysis

In addition to comparing how well students scored on each question, we want to look at which answer options students choose to gain more insight into student reasoning. We determined how many students selected each of the seven answer options (due to the multiple response nature of the question, students could select multiple options, hence we do not expect these numbers to add up to 100%). Table III shows these data with 95% confidence intervals.

For both the PRI and the NRI, students most commonly select B, in which the means of both measurements lie within each other’s error bars. The second most common choices were A and D, in which the error bars of only one of the measurements overlaps with the mean of the other measurement. This shows that, frequently, students require one of the means to be within the error bars of another measurement, as opposed to accepting error bar overlap as agreement between two measurements with uncertainty.

Again from Table III, many more students selected answer option E for the NRI than the PRI; this answer option is the only one where the two measurements definitely do not agree. Students identify this disagreement more frequently when presented with the data pictorially, where it is clear that the error bars are very far from one another, rather than when presented with this same data numerically. During interviews, one student selected all answer options (aside from “None of the above”) on the NRI, and said:

> Honestly I would just say all of them... that’s still at the end of the day what they got... We don’t have enough data to say like ‘no yours are all wrong because they don’t exactly match ours’ because there are a lot of factors that could have altered their numbers and their uncertainty. I know that’s a very idealized way of thinking about science.

However, this student provided expert-like reasoning regarding overlap of the full range of each measurement when correctly answering the PRI, showing a clear difference in thinking about measurement comparison between the two representations.

Knowing the most commonly selected answer options allows us to delve further into common incorrect answer combinations and reasonings for these choices. Figure 2 shows a heat map of the most common answer combinations to each of the two questions (representing 409 of the 670 total student answers). The diagonal represents students who chose the same answer options for both the NRI and the PRI; the off-diagonal elements are students who selected different answers for each of these items.

![Heat map showing the most common answer combinations for the NRI and PRI](image)

One of the more common incorrect combinations on both items is ‘AB’ [NRI: 54/670 = (8±2)%, PRI: 79/670 = (12±2)%]. This incorrect response aligns with students who consider their measurement more important in some sense, and therefore believe that the other groups’ mean must be within their own error bars in order the measurements to agree with one another as compared to the other way around (requiring their mean to be within the other measurement’s error bars), as would be indicated by the selection of ‘BD’ [NRI: 45/670 = (7±2)%, PRI: 63/670 = (9±2)%].

For example, one student who selected only ‘AB’ on the numeric item gave said:

> For the other four groups, the uncertainties for their values did not put them in the same range as my values.
with its uncertainty so I don’t believe they agree with my value.

In other words, when comparing numeric measurements with uncertainty, they placed more weight on their own measurement — in order for agreement to occur, the uncertainty of the other measurement had to encompass their own mean. When solving this problem, they only added and subtracted their uncertainty to their own value and then selected the two answers whose means fell within that range; they ignored the uncertainties in the measurements in the answer options. However, we note that When answering the PRI, this same student selected a correct response of ‘ABCD’, and provided expert-like reasoning. Thus, their reasoning changed with representation.

This theme of placing more importance on their own measurements frequently appeared in student interviews. Occasionally it was present when students provided reasoning for the PRI, but it was more typically found in student answers to the NRI.

Another common incorrect answer for both items is ABD [NRI: 57/670 = (9 ± 2)%; PRI: 60/670 = (10±2)%]. In this line of incorrect reasoning, students did not consider answer option C to be correct, in which just the error bars overlap: they required at least one of the means to be within the error bars of the other measurement in order for agreement between measurements to occur.

Student reasoning from interviews supports this interpretation. For example, one student interviewed selected ‘ABD’ on the PRI because:

Not only do a large portion of their error bars overlap, it also contains the measurement itself,
when referring to answer option B and D. They then chose A because:

I would include [A] because now that measurement is included in mine, but [C] I am not sure about because... I don’t necessarily know for sure they agree.

In this example, the student did not consider answer option C to show agreement despite the error bar overlap - instead, they placed additional emphasis on requiring the mean to be included in at least one of the uncertainties of the other measurement.

Figure 2 also shows that very few students chose ‘ABCDF’ for the PRI [4/670, or (0.60± 0.06)%; but many more students chose this for the NRI [the heat map shows 32 of the 35/670 = (5 ± 2)% students who chose this option]. In the PRI, answer option F is one in which the error bars do not overlap, but are very close to each other, showing that agreement might be possible, hence why selection of F was not considered when scoring this item – this option’s correctness largely depends on which guidelines instructors teach students. Additionally, interview data showed mixed reasoning for students who selected this option.

**IV. CONCLUSIONS & TAKEAWAYS**

Overall, students performed better on the PRI than the NRI, showing a more expert-like understanding of measurement comparison when presented with a pictorial format. However, students’ did not perform as well as desired on either item, indicating room for improvement in teaching this important skill to students. Only about 40% of students correctly identified whether measurements with uncertainties agree with one another in a pictorial format, and this drops to only about 25% when presented numerically instead. Since many scientific papers generally provide numbers with uncertainties for measurements, this is a valuable skill needed in their future scientific careers to interpret experimental results. It is also vital for students to be able to work with many representations of data and convert between them. This study suggests that having students work with multiple representations, and convert between them, could be beneficial for developing expertise with measurement uncertainty and comparing measurements.

In future work, we will examine pre-post gains across this objective by examining scores prior to, and after, instruction in introductory laboratory courses. Additionally, we will explore other research directions using SPRUCE data, such as students’ ideas around accuracy and precision and their ability to propagate errors to obtain an uncertainty in a calculated quantity. Finally, we will examine the alignment of student performance on SPRUCE with a variety of variables, including race, gender, institution type, and instructional methods.

**ACKNOWLEDGMENTS**

This work is supported by NSF DUE 1914840, DUE 1913698, and PHY 1734006. The authors wish to thank Marcos D. Caballero, Rachel Henderson, and Benjamin Pollard for their help in developing SPRUCE, as well as the students and instructors who participated in this study.


An important component of learning physics is being able to apply concepts and reasoning when solving problems. This can be especially challenging for students in quantum mechanics courses, in which the mathematical nature of the theory requires students adjust to new and unfamiliar ways of attaining “understanding”. This paper describes the application of the Essential Skills Framework in an upper-division quantum mechanics course. A preliminary set of “essential skills” were proposed that underlie the solutions of common problems related to probabilities in quantum mechanics. Homework assignments were then developed that provided students with practice in applying these skills. The effect was assessed both by examining the accuracy and speed of students in using these skills over repeated homework assignments as well as through the impact on a standard course exam. Significant improvements were observed, although to a different extent on different types of questions. The results suggest that essential skills practice can be productively incorporated into courses on quantum mechanics, but certain skills are more difficult and may need special attention.
I. INTRODUCTION

Quantum mechanics is recognized as a hard subject; not only are the concepts difficult and unintuitive, but it is expressed in abstract mathematical formalism, much of which is new to students. There has been increasing research examining the problems that students encounter in learning this topic, which ranges from probing their understanding of the many novel physical aspects of the theory [1, 2] to examining their ability to reason with and manipulate the mathematical objects in which the theory is encoded [3].

Ideally, when taking a course on quantum mechanics, students learn both the mathematics and the theory and become adept at translating between them. However, the unfamiliar language and the new ways in which concepts are expressed present a barrier [4, 5]. Even at the start of graduate instruction on quantum mechanics, many students have difficulty in relating mathematical quantities to the relevant concepts (e.g., expectation values, probabilities, eigenvalues, etc.) [6].

One way of interpreting these results is that the combination of new tools and concepts induces considerable cognitive overhead for students [7]. When solving a problem or following an example presented in lecture, students may need to dedicate significant mental energy to simply recall which quantities are represented by different symbols and to distinguish between them. This can impede the ability of students to focus on and recognize the important takeaways of a lesson. It suggests that students need to develop skill in using the novel mathematics (e.g., linear algebra) and learn how to quickly and coherently relate various mathematical representations to real-world phenomena.

In our courses we are using tutorials that are designed to address conceptual difficulties related to quantum mechanics [8]. However, we observe that students continue to struggle with applying the mathematical formalism. We decided to try to give students practice that could improve their fluency in applying the formalism and reduce this cognitive load.

To this end, we designed homework exercises using the online STEM Fluency system based on the Essential Skills Framework (ESF) by Andrew Heckler and Brendon Mikula at Ohio State University (OSU) [9]. It is intended to promote “fluency” by helping students develop both accuracy and speed in applying key concepts and skills. An example in the use of ESF is given by Heckler et al. in the context of vector algebra, which is ubiquitous in introductory sequences and presents difficulty for a wide variety of students [10].

This paper begins by discussing the STEM Fluency system used to provide students with focused practice and gives an overview of some of the skills we have targeted and the questions we have designed. It then presents preliminary results from student responses to a subset of these skills administered over the period of a quarter, as well as a comparison of results from an exam question given in courses with and without essential skills (ES) practice.

II. IMPLEMENTATION AND FORMS OF ASSESSMENT

A. Implementation

Essential skills practice was administered in weekly, required homework assignments via the STEM Fluency system at Ohio State University. This mastery-based, online system allows instructors to create assignments that consist of sets of questions organized into “categories” and “subcategories”. Each assignment includes a set of categories and subcategories chosen by the instructor. The relevant questions administered are then given randomly within the assignment, so sequential questions come from different subcategories. Each category has a “mastery” level, which indicates the number of questions in that category that students must answer correctly in a row before they stop seeing questions from that category in that assignment (thus, having acquired “mastery” in that category). The assignment is complete, and students receive credit, once they acquire mastery in all the categories assigned. Individual students thus saw different numbers of questions based on their performance on prior questions. The individual questions in a given category are all very similar and differ only in the details of the context.

Nine assignments were administered in a junior-level quantum mechanics course ($N = 93$) in Fall 2021. The course used a positions-first textbook (as opposed to a spins-first textbook). On average, each assignment had four categories, each with a mastery level of three. About half of the skills were repeated at least once in the quarter. The assignments were expected to take less than 30 minutes on average. To compensate for the extra work, some homework was removed. All assignments were completed by the majority of the class.

Discussed in this paper is the category we developed for “Probabilities” (offered in Weeks 6 and 7). It had two subcategories Math to Idea and Idea to Math. These were developed by taking into account known student difficulties in quantum mechanics as well as our personal experience in interacting with students [11–13]. It seemed to us that students have difficulty in expressing probabilistic ideas in a mathematical form (Idea to Math) and, conversely, interpreting the mathematical expressions in terms of the relevant concepts (Math to Idea). Note that these questions were designed, to the extent possible, to test single ideas and skills and to be possible to complete quickly.

Figure 1 contains an example question from each subcategory. A full list of the categories and subcategories that were developed, as well as how we decided on the particular question hierarchy, will be discussed in a future article.

B. Forms of Assessment

We have used two methods to assess the impact of the STEM Fluency system on student understanding. The first is by examining the progression of student performance on a
Consider a spin-$\frac{1}{2}$ particle in the state $|\chi\rangle$. The eigenstates for spin in the $z$-direction are given by $|s_i\rangle$, where $s_i = \frac{1}{2}$ or $-\frac{1}{2}$. What do you call/how do you interpret the quantity given by $|\langle \frac{1}{2} | \chi \rangle|^2$?

(A) The probability of measuring the $z$-component of spin for our state to be $\frac{1}{2}$.
(B) The expectation value of the $z$-component of spin in the $|\chi\rangle$ state.
(C) The result of measuring the $z$-component of spin in our state $|\chi\rangle$.
(D) The wavefunction of $|\chi\rangle$ in the $z$-spin basis.
(E) The probability density for $|\chi\rangle$ in the $z$-spin basis.

Let $H$ be a Hamiltonian with energy eigenstates $|n\rangle$ for $n = 1, 2, \ldots$ with associated energy eigenvalues $E_n$. In addition, consider a particle in the state $|\psi\rangle$. What is the probability of measuring $E_2$ for our particle?

(A) $|\langle \psi \rangle|^2$
(B) $|\langle 2 | H | \psi \rangle|^2$
(C) $\langle 2 | \psi \rangle$
(D) $\langle 2 | H | \psi \rangle$

FIG. 1. An example question from a) the Math to Idea subcategory and b) the Idea to Math subcategory of the Probabilities category. Questions in each subcategory had similar structures, involving many different observables and many different representations when available.

![Consider a spin-1/2 particle in the state \( |\chi\rangle \). The eigenstates for spin in the z-direction are given by \( |s_i\rangle \), where \( s_i = \frac{1}{2} \) or \( -\frac{1}{2} \). What do you call/how do you interpret the quantity given by \( |\langle \frac{1}{2} | \chi \rangle|^2 \)?

(A) The probability of measuring the z-component of spin for our state to be \( \frac{1}{2} \).
(B) The expectation value of the z-component of spin in the \( |\chi\rangle \) state.
(C) The result of measuring the z-component of spin in our state \( |\chi\rangle \).
(D) The wavefunction of \( |\chi\rangle \) in the z-spin basis.
(E) The probability density for \( |\chi\rangle \) in the z-spin basis.

Let \( H \) be a Hamiltonian with energy eigenstates \( |n\rangle \) for \( n = 1, 2, \ldots \) with associated energy eigenvalues \( E_n \). In addition, consider a particle in the state \( |\psi\rangle \). What is the probability of measuring \( E_2 \) for our particle?

(A) \( |\langle \psi \rangle|^2 \)
(B) \( |\langle 2 | H | \psi \rangle|^2 \)
(C) \( \langle 2 | \psi \rangle \)
(D) \( \langle 2 | H | \psi \rangle \)

FIG. 1. An example question from a) the Math to Idea subcategory and b) the Idea to Math subcategory of the Probabilities category. Questions in each subcategory had similar structures, involving many different observables and many different representations when available.

given category/subcategory after repeated exposures within STEM Fluency. The second is comparing student performance on exam questions that rely on those skills and that have also been given in prior classes that did not use STEM Fluency.

1. **Assessment within STEM Fluency**

STEM Fluency records each student response and the time taken for each question. We have been using subcategory score (the fraction of correctly answered questions in a given subcategory) as a proxy for “accuracy”. A student who gets 100% accuracy in a given subcategory has not answered any of those questions incorrectly, whereas an accuracy of 50% indicates that the student had multiple incorrect answers. We use time per question (the total time spent on questions in a given subcategory divided by the number of attempted questions in that subcategory) as a proxy for “speed”. Fluency is then a combination of both measures.

For each STEM Fluency assignment, we have found that a few students took a very long time on the assignment. The assignments are designed to be done in less than 30 minutes. Most students take less time, but some take up to 2 hours.

We have dropped these students from the analysis, since they likely took time away from the assignment before finishing and their time is not reflective of the time they spent on individual questions.

2. **Assessment through Exam Questions**

To assess the effect of the ES practice, we analyzed results from two exam questions that were given before (Fall 2018) and after (Fall 2021) ES practice was incorporated into the course. Figure 2 shows the two relevant questions. Both questions can be regarded as testing the Idea to Math subcategory. The content and structure of the course between the two quarters was identical except for the ES practice. Although the instructors differed, our past experience in this and other courses suggests that typically the effect of the instructor is small. Thus, we believe that differences in performance on the question can likely be attributed to the ES practice.

\[
\Psi(x, t = 0) = \begin{cases} 
\frac{\chi_{-\infty}}{a\sqrt{2}}(x^2 - az) & 0 < x < a \\
0 & \text{otherwise}
\end{cases}
\]

You may receive full credit on this problem without evaluating any integrals; however, writing down an integral without explaining where it comes from will not result in credit.

A. Suppose you were to measure the energy of this particle at \( t = 0 \). Determine the probability that the energy is equal to \( E_1 \), the energy of the ground state. Explain your reasoning.

B. Suppose you were to measure the position of this particle at \( t = 0 \) (no other measurements have been made). Determine the probability that the particle is measured to be between \( x = 0 \) and \( x = a/3 \). Explain your reasoning.

FIG. 2. Portions of an exam question administered in classes before and after STEM Fluency was incorporated. The questions assess student ability to determine probabilities for various observables.
through the box-plots in Figure 3. We were interested in determining the impact on all students, thus in each figure the data is divided according to class performance. The terms “Top”, “Middle”, and “Bottom” indicate whether the student’s final course grade was in the top quartile, middle two quartiles, or in the bottom relative to final course grade. The large number of students with perfect accuracy on the second offering motivated the choice of using these three groups.

The “ideal” outcome from using STEM Fluency would be 1) accuracies would increase with a corresponding decrease in TpQ for each group, and 2) lower-performing students would have their final performance near that of the higher-performing students. Here, we focus on two aspects of the box-plots: the medians for each data set (represented by the solid horizontal black lines), and an overview of the distribution of student scores within the Top, Middle, and Bottom groups.

1. Math to Idea Subcategory Results

For the Math to Idea subcategory, the median accuracy increased substantially from the first to the second assignment. All three student groups (Top, Middle, and Bottom) started with a median accuracy between 65% and 85% on the first assignment, which increased to 100% on the second assignment. Thus, half or more of each group, and the class as a whole, performed perfectly with no errors before achieving mastery on the assignment. Interestingly, however, the median TpQ remained roughly the same across all assignments for all three groups. The medians started at 22, 22.4, and 22.9 seconds per question for the Top, Middle, and Bottom groups, respectively, and ended up at 20.8, 21, and 22.1 seconds per question. Thus, after a single repetition, students were completing questions on the Math to Idea subcategory at roughly the same rate, but with much higher accuracy.

Note that in the Top group, essentially every student had 100% accuracy. The bottom halves of the Middle and Bottom groups, despite the large number of students who achieved 100% accuracy, ended with a distribution of accuracies that was broader than it was on the first assignment.

2. Idea to Math Subcategory Results

The Idea to Math subcategory had a different outcome than the Math to Idea subcategory. On the first assignment, performance for the Top, Middle, and Bottom groups was more compressed, ranging from 65% to 70%, so these questions appeared to be harder for the Top group. On the second assignment, only the median accuracy of the top group increased, reaching 100%. The Middle and Bottom groups had mixed improvement. Each had a quarter of their students achieve 100% accuracy and the next highest performing quarter had a higher and broader range than was true for the first administration. However, the median accuracy of the Middle group was unchanged and that of the Bottom group may have had a decrease in median accuracy.

In addition, we saw that the distributions for the lower half of each group broadened considerably. Although many students achieved 100% accuracy on the second assignment, those who did not obtained a wider range of scores compared to those who did not obtain 100% on the first assignment.

Meanwhile, the TpQ of all three groups significantly decreased from the first iteration to the second. The medians started at 20.6, 20.6, and 23.4 seconds per question for the Top, Middle, and Bottom groups, respectively. They ended up at 14, 14.3, and 19.4 seconds per question. Thus, it seems that while the Top group had improvements in both their accuracy and TpQ, the bottom three quartiles of students answered the questions more quickly but with no positive shifts in their median accuracy.

3. Summary and Interpretation

Student performance on the Math to Idea and Idea to Math subcategories demonstrated somewhat different behaviors. The increases in the Math to Idea median accuracies for all groups after only one repetition might suggest that not much practice is required for students to be able to identify the concepts associated with mathematical expressions. The lesser improvement in the Idea to Math subcategory may suggest that translating between probabilistic concepts and mathematics is an asymmetric process, with one direction being more difficult than the other. It suggests that the Middle and Bottom group of students (and perhaps all) may need more
practice in the *Idea to Math* subcategory for them to reach the improvements observed in *Math to Idea* subcategory.

On the other hand, the *Math to Idea* subcategory had no significant shifts in the TpQs from the first to the second assignment. This was true for all groups despite significant improvements in accuracy. The converse was seen in the *Idea to Math* subcategory in which all of the groups had a decrease in TpQ despite the bottom two groups having little or no change in their median accuracies.

One might expect that ES practice would initially result in a simultaneous increase in accuracy and decrease in TpQ and that these would level off with repeated practice. These results suggest that these two aspects do not necessarily improve together; extra practice may be necessary to achieve the desired improvements for both measures.

### B. Assessment Through Exam Question

The essential skills practice was also assessed by using two exam questions given in a class (*N* = 98) without ES practice and in a class with it (*N* = 77)

The questions are shown in Figure 2. Table 1 shows the percentage of students from each class who answered each question correctly. An answer was regarded as “correct” if students had a correct procedure. Questions of this form (short-answer and compound in nature) are different from those used in the ES practice, which had questions that were more atomistic (for example, requiring students to select which bra to use in a probability calculation, or to associate a physics or math concept with a particular mathematical object). The exam question thus requires that students be able to use the skills developed in STEM Fluency when they are not explicitly prompted.

<table>
<thead>
<tr>
<th>Observable</th>
<th>No SF (<em>N</em> = 98)</th>
<th>SF (<em>N</em> = 77)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>43%</td>
<td>66%</td>
</tr>
<tr>
<td>Position</td>
<td>80%</td>
<td>78%</td>
</tr>
</tbody>
</table>

On the energy probability question, about 43% of the students without STEM Fluency answered correctly. This increased to 66% for the class that used STEM Fluency. This 23% increase was statistically significant (with $\chi^2 = 8.547, df = 1$). On the position probability question, both classes did well (80% vs 78% correct), but there was no significant shift ($\chi^2 = 0.062, df = 1$). Given that the only difference was the introduction of STEM Fluency, it is possible that the ES practice may have played a role in improving student performance on the energy question. We believe that this question requires more steps in the solution than that for position and thus STEM Fluency may have played a larger role. However, additional research would be needed to draw conclusions.

### IV. CONCLUSION

Weekly essential skills practice was administered in an upper-division quantum mechanics course. At the start of this project, we identified some skills that we believed research had demonstrated are difficult for students and are essential for students to be able to apply the concepts they were learning. In this paper, we focus on skills related to the concept of probability.

As part of our assessment, we looked at changes in student accuracy and the time required to answer questions within the STEM Fluency system (e.g., performance on repeated assignments). Preliminary results have been encouraging. The scores of students on repeated assignments suggest that students at all levels (Top, Middle, and Bottom, as defined relative to their final course grade) can benefit. There were significant increases in the median accuracies for questions involving translating mathematical expressions in terms of a relevant probability (*Math to Idea*) for each of these three student groups. The results were not as strong for the converse, identifying the mathematical expression corresponding to a given probability (*Idea to Math*). However, even for this type of question we saw considerable improvement, especially among students in the Top group. The disparity in improvement for these two lines of reasoning may not be surprising, since coming up with a mathematical expression requires identifying the component parts (e.g., bras and kets or elements of an integral) and then assembling them. This asymmetry does not appear to be documented in the literature on quantum mechanics and we plan to examine the implications as part of our future work.

We also found that increases in accuracy were not necessarily correlated with the time taken per question by students. This finding suggests that accuracy alone is not sufficient as a measure of fluency. Indeed, learning is a complex process where different aspects may improve at different rates.

Finally, we examined results from exam questions that required students to produce the mathematical expression corresponding to a particular probability (*Idea to Math*). Although not originally designed with STEM Fluency in mind, we saw large improvements on student performance on some of these questions as well.

The preliminary results presented here suggest that incorporating ES practice into a junior level course on quantum mechanics can be productive. This type of practice is not intended to replace other course components that help students develop a conceptual framework. However, the mathematics and formalism in quantum mechanics is sufficiently difficult and new for students that we believe targeted practice in interpreting and generating the mathematical expressions can be useful. In developing our practice questions, we drew on existing research into student difficulties associated with quantum mechanics. However, additional work is needed to determine whether the skills we identified and the questions we designed effectively address the needs of students.
[14] We have evidence from prior quarters that the 2018 results are higher than they would have been, in part, as a result of the use of tutorials in quantum mechanics.
Relating visual attention and learning in an online instructional physics module

Razan Hamed
Department of Physics & Astronomy, Purdue University, 525 Northwestern Ave, West Lafayette, IN, U.S.A.

N. Sanjay Rebello, Jeremy Munsell
Department of Physics & Astronomy, Purdue University, 525 Northwestern Ave, West Lafayette, IN, U.S.A.

Learning using Computer-Assisted Instruction (CAI) demands a high level of attention given the tendency to be distracted and mind-wander. How does the online STEM instructor know when learners are having attentional problems and the extent to which these problems affect learning? In the present study, the visual attentional and cognitive state of physics graduate students were probed while they went through a multimedia instructional module to refresh their knowledge of Newton’s II Law. Data from an eye tracker, webcam, egocentric glasses, screen recording, and mouse and keyboard events were integrated to record learners’ attention overt attention to the learning environment (+/-) and thinking about learning content (+/-) to analyze students’ attention spans during learning from this module. On average, learners were found to be on-task and on-screen for a vast majority of time, with evidence of mind wandering. The learning module improved the participants efficiency with which they answered the questions correctly on a post-test relative to the pre-test. Further, there is a positive albeit statistically non-significant correlation between the improvement from pre- to post-test efficiency and the time spent on-screen and on-task during the module.
I. INTRODUCTION

Computer-Assisted Instruction (CAI) is ubiquitous and will remain so in the post COVID-19 world. A key issue with CAI is students’ attention during instruction. The causal relation between attention and learning is well established. When learners ignore relevant information, or attend to irrelevant information, it reduces learning. [1–4] To address this issue, learners’ attention is cued to relevant information, or remove irrelevant information, in the CAI materials. [1–4] But such solutions are only effective if learners attend to the CAI materials in the first place! Lack of sustained attention to CAI learning materials reduces learning. [5,6] However, researchers have only begun to operationalize what is meant by “students’ attention span.” Research has shown that a student may be looking at CAI materials on their computer screen, but not thinking about them, because they are mind-wandering. [7–9] Alternatively, a student may be looking away from their computer screen, but thinking about their learning materials, such as while note taking or reflecting. So, we must clarify the notion of attention span itself in the context of CAI environments.

II. THEORETICAL FRAMEWORK

Research has demonstrated a way of encapsulating all these different factors that affect learning [7]. The first factor is attention: being generally inattentive vs. generally attentive to the learning environment, which can be measured by whether students are looking at the learning environment. The second factor is thinking: thinking about the learning materials vs. thinking about something else (e.g., mind wandering). The proposed research builds on D’Mello’s [7] 2x2 matrix to characterize learners’ attentional states during educational activities. In this framework, learners can transition between four attentional states that consider both the overt and covert aspects of attention (Fig. 1).

**Quadrant 1 (Q1) Top Left:** Learner visually attends to the learning environment (on-screen), while thinking about it (i.e., on-task).

**Quadrant 2 (Q2) Top Right:** Learner does not visually attend to the learning environment (off-screen) but thinks about it (on-task) e.g., note taking or using the calculator to solve a relevant problem.

**Quadrant 3 (Q3) Bottom Left:** Learner visually attends to the learning environment (on-screen) but does not think about it (off-task) e.g., mind wandering.

**Quadrant 4 (Q4) Bottom Right:** Learner neither visually attends to the learning environment (off-screen) nor thinks about it (off-task) e.g., distracted by a cell phone or text.

<table>
<thead>
<tr>
<th>Content-related thoughts</th>
<th>Overt attention to learning materials (computer)</th>
<th>Overt attention elsewhere</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quadrant 1 (Q1)</td>
<td>Overt sustained attention to one or more areas of the learning materials</td>
<td>Quadrant 2 (Q2)</td>
</tr>
<tr>
<td>Covert sustained attention (e.g., note-taking, using calculator)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quadrant 3 (Q3)</td>
<td>Covert inattention (mind wandering)</td>
<td>Quadrant 4 (Q4)</td>
</tr>
<tr>
<td>Overt inattention (e.g., off-task, distracted)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FIG. 1. The 2x2 attentional-cognition matrix.

III. RESEARCH DESIGN

A. Research Questions

We addressed the following research questions. RQ1: What percent of time during the instructional module did participants spend in the four attentional-cognitive states (Q1, Q2, Q3 and Q4) defined by overt attention to the learning environment (+/−), and thinking about learning content (+/−)? RQ2: How does the percentage of time that participants spent in Q1 correlate with their learning from the module?

B. Participants

The (N=12) participants for this study were recruited from a pool of students enrolled in the physics graduate program at large U.S. midwestern land grant university. All the participants had a bachelor’s degree in physics or an allied science discipline and therefore had been previously exposed, during their introductory undergraduate education, to the material covered in the instructional module. As participants had been exposed to this learning material previously in their academic preparation, this module served as a refresher of the learning material for the participants.

C. Materials

The materials consisted of a pre-test, module, and post-test. The multimedia module was about 15 minutes long and was designed using the backward design strategy of [10] and consistent with [11] principles of multimedia learning. The module focused on reviewing Newton’s II Law, free-body diagrams, and solving problems based on these concepts. During the module each participant was presented with a “mind-wandering” prompt in which they were explicitly asked to respond on the keyboard with a “Y” (yes) or “N” (no) to indicate whether they were mind-wandering. The mind-wandering prompts appeared randomly and continually throughout the instructional module, excluding the pre-test and post-test. The time duration between each prompt and the next was within a range of 120 seconds to 240 seconds. The pre-test and post-test each had seven multiple choice items, that included four conceptual questions and three problems that required an application of the problem-solving strategies presented in the module as
well as use of scratch paper and a calculator, which were provided to the participants. In addition to the module and the tests, the study utilized annotation software which flagged instances when the participants eyes were off-screen. The software also allowed participants to view their recording of those instances in a retrospective recall interview and self-identify, with assistance from the research assistant, which attentional-cognition state (Q1, Q2, Q3, or Q4) they were in during those instances.

**FIG. 2. Study design.**

D. **Data collection process**

Figure 2 shows the design of this study which took a total of 70-80 minutes for each session. During the process (pre-test, module, post-test), data from various sources (See Fig. 3) were collected: webcam, egocentric camera, screen recording, mouse and keyboard events, and eye-tracker). All sources were synchronized and recorded by specially designed software. Upon the completion of each recording, the software identified segments of time when the participant’s eyes were off screen for a duration between 3-12 seconds. For each such segment of time, the software created a video clip and stored each segment as a series of 9 still frames or images.

After completion of the post-test, each participant completed a retrospective recall interview which lasted about 15-20 minutes. During this interview they were presented with video clips (each shown as a series of nine still frames) showing their screen, face (recorded from the webcam) and egocentric view (recorded from the egocentric camera). They were asked to reflect on that instant of time shown by the series of still frames and identify whether they were on/off-task and whether they were looking on/off screen (although the eye-tracker is only supposed to flag them when they are looking off-screen, sometimes it may be unable to track their eyes even when they are looking on-screen.)

**FIG. 3. Data collection sources. In the instant shown the participant is taking notes, which would be coded as off-screen/on-task (Q2).**

IV. **DATA ANALYSIS**

For the duration when the eye-tracker had flagged the participant as off-screen, we used data from the participants’ self-reports in the retrospective recall interviews to categorize the time segment in either Q2 or Q4. For the duration when the eye-tracker had not flagged a time segment as off-screen, the participant by default was deemed to be on-screen and on-task (i.e., in Q1) unless they had responded “Y” to the mind-wandering prompt. In that case, they were deemed as mind-wandering (i.e., in Q3) for half the time between the previous prompt and the current prompt. The data from the participant self-reports was collected during the retrospective recall interview with data from participant responses to the mind-wandering prompts presented during the module.
V. FINDINGS

1. Duration in Attention-Cognition States

Table 1 shows the time spent by the participants in each quadrant. On average, participants spent a vast majority (85%) of their time in Q1 (on-screen and on-task). This indicates that the participants were highly engaged in the task for a vast majority of the time. This is expected given that the participants were mature graduate students who were less likely to be distracted from the task. However, we also found that on average participants were spending about 10% of their time in Q3 (on screen and off-task), which indicates that they were mind wandering about 10% of the time. This too is expected given that the participants who were graduate students were familiar with the content and therefore may have found the content boring therefore were more susceptible to mind wandering than say participants such as undergraduates who were more likely to be unfamiliar with the content and may have been more interested and motivated in learning the content. Finally, we also find that participants almost a negligible percentage of their time in Q4 (off-screen and off-task), perhaps because they were completing this task in a research lab rather than in a naturalistic environment in which case off-screen and off-task behaviors might occur more frequently.

<table>
<thead>
<tr>
<th>Quadrant</th>
<th>Average Time in Quadrant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>820.6 ± 170.7 s</td>
</tr>
<tr>
<td>Q2</td>
<td>63.2 ± 124.4</td>
</tr>
<tr>
<td>Q3</td>
<td>72.4 ± 101.7</td>
</tr>
<tr>
<td>Q4</td>
<td>2.0 ± 6.6</td>
</tr>
</tbody>
</table>

We determined the gain in the score/time metric for each participant i.e., the change from pre-test to post-test, as given by equation. We wanted to test whether the change in Score/Time metric correlated with the time spent by learners in Quadrant Q1 (on-screen and on-task) as spending time in this quadrant is expected to be most conducive to learning. We found a positive, but statistically insignificant correlation coefficient (0.32) between the Gain Score/Time and the time (in seconds) spent in Quadrant Q1. (Fig. 4). Even though the result is not statistically significant, it may suggest a trend that participants who spent more time on-screen, and on-task learned more from the module, which is an expected result.

2. Relation between Attention-Cognition and Performance

The pre-test and post-test showed a ceiling effect, which is expected because the participants were graduate students. However, an interesting metric in this case is the efficiency in providing the correct answer to pre-test and post-test items. This metric is defined as SCORE/TIME or S/T and it is calculated for each participant on the pre-test and post-test using equation (1):

\[
\frac{S}{T} = \sum_{i=1}^{M} \frac{s_i}{t_i}
\]

where \(s_i\) is the score (0 or 1) on item \(i\), and \(t_i\) is the time in seconds taken by the participant on item \(i\), where \(i\) ranges from 1 to \(M\), where \(M\) is the total number of items on the assessment, which in this case \(M = 7\).

We argue that SCORE/TIME metric is a relevant metric for high prior knowledge participants where the pre-test and post-test scores show the ceiling effect, because this metric considers not just whether the participant answered an item correctly also it considers the time spent to answer the items. Because the time spent in the denominator, the greater the time taken the smaller the contribution to the SCORE/TIME metric for the participant. In other words, the SCORE/TIME metric rewards participants if they answer an item correctly but penalizes participants if they take a longer time to answer the item.

IV. CONCLUSIONS

To address our research questions, we find that participants spent a vast majority of their time in Q1 (on-screen and on-task), but they also spent a significant time in Q3 (on-screen and off-task) i.e., mind wandering. This could be attributed to the fact that the participants were using the module to refresh their knowledge in this area and therefore may have been bored during some aspects of the module.

A ceiling effect on the pre-test scores for these participants, given their high level of prior knowledge of the module content since they were all graduate students in physics and the module was targeted at the introductory undergraduate physics level. However, the speed with which the participants were able to answer these questions improved such that their efficiency in answering the questions correctly was improved from pre-test to post-test. Moreover, the increase in efficiency in answering the questions correctly was improved from pre-test to post-test. We also found that the increase in efficiency in answering
the post-test questions correctly was positively correlated, albeit not statistically significantly, with the time they spent in Q1 i.e., on-screen and on-task.

V. LIMITATIONS & FUTURE WORK

There are several limitations to this study that will be addressed in future work. First, the participants were high prior knowledge learners and showed a ceiling effect on the pre-test. They also were less likely to be off task and might have experienced a practice effect on the post test. We plan to repeat the study with low prior knowledge learners who might not show such an effect and show a greater propensity to be off task. Second, the study was not done in a naturalistic setting i.e., the participants were not at home or other locations where they are more likely to be distracted and mind-wander. Third, the students self-reported either being focused or mind wandering when answering the mind wandering prompts which is not a reliable way of categorizing the time that they spent looking off the screen. In future, we plan on using more objective prompts rather than using yes or no questions to detect mind wandering.

Research has shown that attention plays a critical role in online learning. However, researchers have only begun to operationally define attention or connect students’ attentional states to their learning outcomes. This study is the first step to bridge the theory and methods of studying attention in cognitive science with educational practice. It points the way to making progress in understanding the connections between students’ moment-to-moment attentional states and their STEM learning. It is a first step to deepen our understanding of attention/learning processes.

This study contributes to research on online learning. We have adapted and operationalized a theoretical framework to measure the moment-by-moment attention-cognition states of learners completing an online module and explore the relationship between these states and their learning outcomes on the module.

ACKNOWLEDGMENTS

Supported in part by U.S. National Science Foundation grant DRL-2100218. We acknowledge the contributions of our collaborators who assisted with the design of the data collection software and analysis of the eye movement data.
A Look Physics Teacher Identity Around Equitable Instruction: The Tour Guide, Coach, and Gardener

Maria S. Horak\textsuperscript{1} (she/her), Clausell Mathis (he/him)
\textsuperscript{1}Lyman Briggs, Michigan State University, 919 East Shaw Lane, East Lansing, MI, 48823

Delwrick Nanthou\textsuperscript{1} (he/him), Abigail R. Daane\textsuperscript{1} (she/her)
\textsuperscript{1}Department of Physics, South Seattle College, 6000 16th Ave. SW, Seattle, WA, 98106

Michelle N. Brown\textsuperscript{2} (she/her)
\textsuperscript{2}Curriculum and Instruction, Pennsylvania State University, 144 Chambers Building, University Park, PA 16802

Studies show that a physics teacher's identity impacts how they teach. For teachers interested in pursuing diversity, equity, and inclusion practices in physics instruction, their identity likely informs how they conceptualize equitable instruction. We highlight three teacher identity cases from a data set of 25 interviews of secondary and university physics teachers to examine how they conceptualized equitable instruction through four domains: their conceptions of self, others, knowledge, and pedagogy. Selected teachers had distinct conceptualizations around enacting equitable instruction that we described across three metaphors: a tour guide, a coach, and a gardener. This study urges researchers to be more attentive to the varying dynamics of physics teacher identity and its impact on teacher practice.

Keywords: physics teacher identity, equity, teacher conceptions
I. INTRODUCTION

A teacher's identity influences practice and how they embrace reform efforts [1, 2]. This is why identity has been researched with the goal of helping teachers maximize effectiveness [2, 3, 4]. An examination of the literature on teacher identity [2, 5, 6, 7] shows several consistent themes: (1) teacher identity is socially constructed; (2) teacher identity is dynamic; and (3) teacher identity is complex and multivariate, consisting of intersecting sub-identities. These sub-identities of teachers include: how they position themselves with a group; how they view what they teach; how they interact with students; how they feel about others; and how others feel about them. Also, other identity markers such as race, gender, and sexuality impact how teachers view themselves and their interactions with others. Within physics education, research on teacher identity is scant. Teacher identity has been primarily researched in social science disciplines, with a small amount in STEM disciplines. This paper will describe a study on the physics teacher identity of three teachers to understand the differences and dynamics of their conceptions of self, others, knowledge, and pedagogy. We expect our study to initiate the expansion of research on physics on teacher identity by introducing a new analytical lens on physics teacher identity toward equitable instruction.

II. ANALYTICAL FRAMEWORK

A. Physics Teacher Identity

Physics teacher identity refers to how one sees oneself as a physics teacher [8]. Teacher identity is important because it provides a lens to examine teachers' perspectives on “how to be, how to act, and how they view their place in society” [9]. Teacher identity is developed based on many factors, such as experiences, values, and interests. Identity is dynamic and can shift over time based on the influences of many internal and external factors [10].

In an analysis of focus group interviews of physics teachers who participated in a professional development workshop on equitable instruction [11], we found teachers had various perspectives on how equitable instruction is actualized. Taking inspiration both from this analysis and the culturally relevant theoretical framework [12, 16, 17, 18], we developed a new framework for physics teacher identity towards equitable instruction that consists of four domains: conceptions of self, others, knowledge, and pedagogy.

Conceptions of self focus on how physics teachers view themselves and their ability to teach for equity. A teacher's conception of self on equitable instruction can consist of their self-efficacy; pedagogical discontentment; their belief in the ability to: connect, motivate, inspire, and relate to students; and their ability to reconcile their experiences with their beliefs towards their teaching.

Conceptions of others focus on physics teachers' view of their students. This includes teachers' views about students' abilities, including those related to cognitive processing of physics ideas, doing scientific and/or academic tasks, and epistemic strengths.

Conceptions of knowledge focus on physics teachers' views of canonical (content) knowledge. This can be examined in teachers' belief in objectivity in physics or their ability and/or willingness to problematize physics theories and ideas.

Conceptions of pedagogy focus on how teachers view their ability to teach for equity. This includes views of how equitable instruction is actualized through teachers' moves, student engagement, lesson planning, assessment design, and integrating content and scientific practices.

III. RESEARCH QUESTIONS

Conceptions about equitable instruction vary among teachers. The goal of our study is to unpack these different conceptions of how physics teachers develop equitable instruction. To do so, we answer two research questions: What are the different conceptions of physics teacher identity across self, others, pedagogy, and knowledge? How do teachers' definitions of equity intersect with their physics teacher identity?

IV. METHODS

A. Data Collection

A group of physics teachers (n=25) was selected for interviews based on demographics, years of experience, and gender. The teachers were interviewed through a semi-structured interview protocol that asked about their conceptions of self, others, knowledge, and pedagogy. Some interview questions were: “If you could describe your teaching identity, what would it be? Why do you think some students struggle/fail? Do you think physics knowledge is objective and free from bias? What is the best approach for teaching equitable instruction? Why?” We selected these teachers based on educational role, race, gender, years of profession, and school demographics. The interviews were held online where they averaged around 60 minutes in length.

B. Data Analysis

We reviewed the data to identify similarities and differences across teachers. Using a multiple case study approach [13], we identified three teachers who exemplified different identity patterns to better conceptualize different constellations of physics teacher identity and their relationship to teaching for equity. Each
teacher was given a pseudonym: Leonard, Layla, and Ruby. We coded and analyzed videos using online software (Vosica). Nvivo was used to analyze transcripts of teachers through thematic coding over several cycles. In our first coding cycle, we identified common conceptions and grouped them around self, others, knowledge, and pedagogy. This allowed us to identify common patterns across identities, which we applied to a second coding cycle that conceptualized different groupings of identity. Through the second coding cycle, we identified the different types of identity markers at a smaller grain across teachers' conceptions of self, others, knowledge, and pedagogy. For example, a teacher's conception of self could be one who values self-improvement. We later mapped these onto the metaphors of tour guide, coach, and gardener.

III. FINDINGS

After analyzing interview data, we found that teacher responses aligned with three metaphors for teaching identities: tour guide, coach, and gardener. Although we recognize that these groupings are not completely disparate, they reveal a pattern of types of identity along with definitions of equity. In the sections below, we introduce each teacher's case and their self-identified metaphor to better characterize different constellations of teacher identity.

A. Leonard the Tour Guide

Leonard is a white male university physics instructor with 15 years of teaching experience. He describes himself as a "tour guide who is on a journey with students, helping them navigate a path to expand their physics knowledge." Through this metaphor, Leonard sees himself as a teacher who helps visitors (students) navigate a terrain (physics knowledge). However, Leonard also shared that he is not perfect, and although he thought equitable teaching was an obtainable goal, he did not believe he had mastered it yet.

Leonard’s tour guide analogy aligns with his conceptions of pedagogy. When asked to describe his teaching approach, Leonard described it as more lecture oriented, focused on explaining physics concepts and showing students skills for problem-solving. However, Leonard also shared teaching strategies that were more student-centered. When asked about teaching for equity, Leonard responded that he used strategies such as getting students to talk to each other through think-pair-share, working in groups to do quizzes and problem sets, and having class discussions around physics topics. Although Leonard’s pedagogy was diverse, it largely aligned with a traditional didactic, teacher-as-knower stance, which positions students as blank slates [14].

Leonard believed all physics knowledge was correct and made of facts, adding: “The stuff I teach is correct physics as far as humanity knows it today.” This corresponds with a traditional tour guide stance that shares facts which are posited as truth rather than positioning knowledge as subjective or biased.

Leonard stated that he wants his students to succeed, “however, they have a responsibility for their own learning.” Leonard explained further:

Students have responsibility for their own education. An unwilling student that’s not motivated [to learn], that’s on them, and it [the responsibility] should not be removed from them and given to the institution or the teacher.

When looking at this response through a tour guide metaphor, we can similarly find that tourists are responsible for what they take away from their guide’s presentation.

Within the context of schooling, Leonard’s response also speaks to his definition of equity. When Leonard was initially asked if he wanted to teach for equity, he responded in the affirmative, adding that it can be done in many ways if there is active learning and participation among students.

When he was later asked: “What is equitable student participation?” Leonard appeared confused, stating, “[I’m] not sure how to interpret that question. I don't know what equitable student participation is.” Leonard further reflected: “I think we should have opportunities for every student to participate if that's what you mean by equitable, fine. But that's not equitable participation; that's equitable opportunities.” Leonard said he expected all students to “be vocal in class” and that “every student gets a chance to answer questions.” This stance was reiterated when Leonard agreed with the interviewer’s definition of equity, “students having the support and resources they need to thrive as whole humans in school–including academically, socially, and emotionally,” however he added that he is for “equal opportunity and participation, but that does not mean equal outcomes.” Much like a tour, the opportunity is the same for each visitor.

B. Layla the Coach

Layla is a white female high school physics teacher with ten years of teaching experience. Layla metaphorically described her teaching identity as a coach. She stated she wants to support and “coach” [students] through the learning process.

Her identity as a coach mapped onto how she thought of her students and of teaching. Layla stated that she wants students to feel comfortable within the classroom. She described pushing her students to embrace “productive struggle” in understanding physics. When thinking about pedagogy, Layla discussed readjusting the classroom to meet the individual needs of her students. She reflected:

I think that there are certain groups of people that in a science classroom can come [in] and automatically feel like they [...] play an important role in the science world. And then some people that feel like they have not been told that they play an important role. And so
think part of equitable teaching is making sure that everybody feels like they have an important role to play, and so maybe not just support it in their learning of the material, but also support it in their identity as a scientist.

Later in the interview, Layla stated, “I’m not trying to convince everyone to be a scientist, but I want them in the science classroom to feel like they are a scientist.” Layla’s statements on wanting students to “feel” as if they are a scientist indicate a teaching focus on strengthening students’ perceptions of their capability of doing physics. In addition, Layla stated that she adjusts her teaching to the needs of the students, both individually and as a group, because students have different needs for learning. As a coach, Layla described wanting to foster improvement in each student and continually looking for better tactics to do so.

Regarding the discipline of physics, Layla described physics knowledge as a model to represent how the world works. Notably, she recognized the limitations of models due to human error and bias. Layla also viewed physics knowledge as a tool to understand and apply phenomena.

Layla described equity as related to respecting students and setting them up for success:

- I definitely think [equity is] attainable in the sense that we can be at a place where we’re like: We are now honoring our students and supporting them and giving them the resources they need. For sure, but I think they’ll always be something you can do better.

Layla’s coach metaphor permeated her conceptions of self, other, and pedagogy. In addition, unlike Leonard, she recognized the sociocultural influence of knowledge and considered students’ individual needs when discussing equity.

**C. Ruby the Gardener**

Ruby is a white female physics teacher at a two-year college with 21 years of teaching experience. In describing her teaching identity, Ruby referred to herself as a gardener:

- I care about [my students] as humans growing, as opposed to thinking of myself as some sort of test-giver...I try to foster growth in the students. I’m thinking of a garden: instead of weeding out students, I'm trying to have each little seedling grow.

In her conception of others, Ruby aligned with the gardener metaphor by recognizing that her students are diverse, each with their own experiences and perspectives. Ruby shared that many of these outside experiences cannot be controlled by the teacher and, therefore, can make it impossible to teach thoroughly for equity without changing those external systemic issues.

The garden metaphor also applies to Ruby’s conceptions of pedagogy. Ruby recognized that learning is not just about imparting knowledge but also about cultivating an environment that supports student growth and development.

She stated that she preferred student learning to be a community effort and not an individual task.

The importance of perspectives emerged in Ruby’s conception of knowledge, where she posited physics knowledge as biased and cultural:

- Culture comes in [to] almost every aspect [of physics knowledge]: our own internal bias, confirmation bias, all of the biases that we have when we're interpreting data. [...] Who gets to ask the research question? Who gets to focus the research? Who gets the money to ask their own research question? Who gets hired in the first place? Who even gets an interview?

Similar to Layla, Ruby understands the role of bias and its impacts on research and knowledge. However, she adds more complexity and criticality when describing the role of bias in physics knowledge creation. Ruby continues:

- There is another set of barriers that are specifically aligned with the dominant culture, the white culture, and even the naming of our laws. Why are we naming them after people and not naming them for what they are [about]? [...] It’s because the people in power are the white males, and so every single physics law is after a white male. It’s amazing how illogical that is, even though you know there were Muslim men who actually discovered things before the white males that have their names on things. Even Ibn Sahl, for example [...] that should not be Snell’s Law, it should be Ibn Sahl. So it’s all Eurocentric, and every step of the way is biased...Oh, yeah, is it subjective? Absolutely.

Here Ruby explicitly names whiteness and provides a critical perspective that considers power asymmetries in regard to physics contributions. She names examples where white, European men’s names have been attached to concepts with non-European roots. This differs significantly from Leonard’s conception of objectivity in physics and a meritocratic definition of equity, as well as from Layla’s definition of bias as tied to human error and a more individualistic description of equity. Ruby’s statement indicates that she views the bias within physics as impacting what students learn. She acknowledges how the traditional physics canon is deemed credible and does not consider the different perspectives or contributions of individuals from non-Eurocentric communities.

Similar to her conceptions of knowledge, Ruby viewed issues of equity as systemic and larger than the classroom. She shared that outside factors, such as socioeconomic status, family background, and access to resources, can significantly impact a student’s ability to learn and succeed. Despite these challenges, Ruby shared a desire to provide a supportive and inclusive learning environment that helps all students to achieve their full potential. As such, Ruby sought to create a learning environment that meets students where they are and provides them with the resources and support they need to succeed. Ruby’s conceptions of her own role as a teacher, her students, teaching, and knowledge map to her gardener metaphor, in that she sees...
IV. DISCUSSION & CONCLUSION

Three metaphors emerged from our study to help us articulate some of the nuances of physics teacher identity across self, others, pedagogy, and knowledge. In addition, considering teachers’ definitions of equity alongside these metaphors helps us consider the relationships among them.

A. Productivity of the Metaphors

The metaphors from the study help us better understand the interconnections across conceptions of identity. For example, implicit in each metaphor is a representation of one’s perspective of self as a teacher: Leonard’s tour guide metaphor aligned with taking students on a journey of learning; Layla’s coach metaphor positions her as helping students in their learning; and Ruby’s gardener metaphor situates her as wanting to support students in their growth.

These metaphors map onto pedagogical stances, with the tour guide representing a lecture-centric identity that positions the teacher as all-knowing. In contrast, the coach and gardener metaphors map to a student-centered practice, where teachers support student learning through scaffolding and feedback (coach) or providing the resources needed to thrive (gardener). The gardener metaphor also considers the classroom context, critiquing how cultural and systemic norms operate.

When viewing the three metaphors through teachers’ conceptions of knowledge, we can also identify three different paradigms. Leonard believed physics knowledge was objective, aligning with a post-positivist lens. Layla recognized the utility of physics in modeling the world. She also noted how physics is created by humans and is rife with bias and limitations, aligning with a constructivist paradigm. Ruby discussed how physics knowledge was created in biased and Eurocentric systems, aligning with a constructivist and sociocultural paradigm along with a critical examination of power dynamics.

The metaphors help us understand how the different aspects of identity intersect. Leonard’s tour guide metaphor conceptualizes a traditional stance that positions students as blank slates and teachers as knowers, with lectures as a central practice. In this stance, students are responsible for their own learning and success, and physics knowledge is positioned as an objective fact. Layla’s coach metaphor conjures a constructivist and student-centered practice, where students are recognized as knowers with individual strengths, and physics models are tools that may include implicit bias. Ruby’s garden metaphor recognizes the role external environments and systems have on education, critically looking at the role of power dynamics in how knowledge is created and taught. Although we recognize that no teacher, even the three cases presented here, fit neatly into one category, the metaphors provide useful tools to critically examine the constellations of physics teachers’ identity across different domains.

B. Defining Equity Across Metaphors

Teachers’ understandings of equity significantly impact how they identify and describe their teaching. Leonard’s definition of equity was closely aligned with equality. For example, Leonard stated that students fail to succeed as a result of their own poor work ethic, disregarding how systems of oppression play a role in dis/advantage and aligning with the myth of meritocracy [15].

Layla defined equity as providing students with the resources they need to be successful. In addition, she was aware of students’ diverse identities and wanted to make sure students felt represented in the classroom. Layla’s definition aligned with her metaphor of coach, where she believed in meeting students where they are and helping them improve. Unlike Leonard, she positioned student failure in her class as due to external factors, such as the need for more time with students. This belief aligned with other evidence of her self-reflective nature and her desire to improve.

Lastly, Ruby’s definition of equity recognized how each student has unique strengths and needs and how outside factors play a role in teaching equitably. Unlike the other teachers, Ruby discussed systemic aspects of inequity. With a systematic perspective, Ruby acknowledged that she would never be able to become a completely equitable teacher but would strive to constantly improve. This perspective acknowledged outside factors, similar to Layla’s concern with time, as playing significant roles in student success.

Although we recognize the power of using these metaphors to label teachers’ identities and the relationship between identities and conceptions of equity, we also found identities to be fluid and non-binary. For example, besides didactic teaching examples, Leonard also shared student-centered practices, such as group work. We share these metaphors and their correlating definitions of equity to better conceptualize physics teachers’ identities. We do not assume physics teachers’ identities are static or neatly aligned into any one of the metaphors. Our intention is that these metaphors and teachers’ varying definitions of equity provide helpful conceptual tools to frame future work in this space.

ACKNOWLEDGMENTS

We gratefully acknowledge all the teachers in the study for their generosity in allowing us to interview them and share their conceptions that were accessible to the research team. We are grateful to the [Anonymized Research Group] at [Anonymized University] for their support in critiquing this work.
Development, validation and online and in-person implementation of clicker question sequence on change of basis

Peter Hu, Yangqiuting Li, and Chandralekha Singh

Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15260

Research-validated multiple-choice questions comprise an easy-to-implement instructional tool for scaffolding student learning and providing formative assessment of students’ knowledge. We present findings from the implementation of a research-validated multiple-choice question sequence on concepts relating to and various methods of changing basis for two-state systems. This study was conducted in an advanced undergraduate quantum mechanics course, in both online and in-person learning environments, across three years. Student learning was assessed after traditional lecture-based instruction in relevant concepts, and students’ performance was compared with that on a similar assessment given after engaging with the multiple-choice question sequence. We analyze, compare, and discuss the trends observed in the three implementations.
I. INTRODUCTION

Two-state systems are often used to illustrate many rich phenomena in quantum mechanics (QM), due to their relative simplicity compared to higher dimensional Hilbert spaces. Furthermore, since we are in the midst of the second quantum revolution [1,2], they are critical to the field of quantum information science to describe the behavior of qubits, the smallest unit in which quantum information is stored and processed. Yet because of unfamiliarity with the quantum formalism, even advanced undergraduate students can often struggle with concepts such as basis and changing the basis of two-state quantum systems and knowledge relevant to these concepts. Here, we present research regarding instruction on these concepts, namely outer products and changing between the $z$-basis and $x$-basis, two common bases in two-state systems that use the eigenstates of the $\sigma_z$ and $\sigma_x$ Pauli matrices, respectively. Outer products are crucial in understanding and being able to carry out a basis change by applying the identity operator expressed in the appropriate basis. The importance of basis and change of basis in two-state systems has been emphasized in prior work [3,4].

Prior research suggests that students in QM courses often struggle with many common difficulties [5–21], including change of basis. For such difficulties, research-validated learning tools can effectively help students develop a robust knowledge structure [22–27]. For example, tools known as Quantum Interactive Learning Tutorials (QuILTs) have been developed with encouraging results on many topics in QM [28–30]. Other commonly used learning tools in physics include clicker questions, first popularized by Mazur [31] using the Peer Instruction method. These are conceptual multiple-choice questions presented to a class for students to answer anonymously, typically individually first and again after discussion with peers, and with immediate feedback to instructors using an electronic response system, generally referred to as clickers [31]. They have proven effective and are relatively easy to incorporate into a typical course, without the need to greatly restructure classroom activity or assignments [32].

While these multiple-choice questions can be successfully implemented in physics classrooms without additional technological tools, this research used clickers, which automatically tracked student responses in real time. When presented in sequences of validated questions on a particular topic, clicker questions can systematically help students with particular concepts that they may be struggling with. Previously, such multiple-choice question sequences, or Clicker Question Sequences (CQS), related to several key QM concepts have been developed, validated and implemented [33–36]. Here, we describe the development, validation, and implementation of a CQS intended to help students learn about outer products and change of basis in two-state quantum systems.

II. METHODOLOGY

The developed, validated and implemented CQS is intended for use in upper-level introductory QM courses. The data presented here are from administration in a mandatory first-semester junior/senior-level QM course at a large research university in the United States. During the development and validation process, we took inspiration from a previously-validated QuILT and CQS on Dirac notation, since much research involving cognitive task analysis had been conducted from both student and expert perspectives [28]. We also drafted and iterated new questions for the basics of two-state systems, including outer products and change of basis. To ensure that the material could be completed in the allotted class time, while offering maximal value to students, we prioritized the coverage of conceptual knowledge, used common difficulties as a guide, and provided checkpoints that could stimulate useful class discussions. We iterated the questions many times amongst ourselves and with other faculty members to maximize clarity, consistency, and accuracy. We then conducted think-aloud interviews with five students in order to fine-tune the questions and ensure that they are at an appropriate level.

The learning objectives of the CQS are threefold. CQS 1.1-1.4 help students identify properties of two-state spin systems and spin-1/2 systems in particular. CQS 2.1-2.6 help students achieve fluency in translating between Dirac notation and matrix representation and calculating outer products. Finally, CQS 3.1-3.5 build on this prior knowledge to help students change the basis of a quantum state through several approaches. In particular, the three methods discussed to help students be able to change basis were direct substitution (e.g., using $|\pm x\rangle = \frac{1}{\sqrt{2}} (|+z\rangle \pm |-z\rangle)$ to replace a state expressed in the $x$-basis using the standard notation); the viewing of inner products, e.g., $\langle \pm x| x\rangle$, as projections of the state $|x\rangle$ along the basis vectors in a particular basis (in this case, the $x$-basis); and the use of spectral decomposition of unity. Here, we report on the results of selected concepts, namely outer products and change of basis, for which students had their difficulties greatly reduced. The clicker questions relevant to these concepts are provided below.

CQS 2.2 Given that

$|+x\rangle = \frac{1}{\sqrt{2}} (|+z\rangle + |-z\rangle),$

$|-x\rangle = \frac{1}{\sqrt{2}} (|+z\rangle - |-z\rangle),$

$|+y\rangle = \frac{1}{\sqrt{2}} (|+z\rangle + i|-z\rangle),$

and $|-y\rangle = \frac{1}{\sqrt{2}} (|+z\rangle - i|-z\rangle),$

choose all of the following statements that are true:

I. $\langle +z|+z\rangle = 1$ and $\langle -x|+x\rangle = 0$

II. $\langle +y|+y\rangle = 1$ and $\langle +z|+y\rangle = \frac{1}{\sqrt{2}}$
III. \( (+x) + x = 1 \) and \( (+x) - y = \frac{1}{2}(1 - i) \)
A. I only B. I and II only C. I and III only D. All of the above E. None of the above

CQS 2.5 Use the following matrix representations:

\[ |+z\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad |-z\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \]

\[ |+x\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad |-x\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \]

Choose all of the following statements that are true:

I. \( |+xz\rangle = \frac{1}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \)
II. \( |+xz\rangle = \frac{1}{2}(1 - 1 + 1 - 1) = 0 \)
III. \( |+z\rangle + |z\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \)

A. I only B. II only C. III only D. I and III only E. None of the above

CQS 2.6 Use the following matrix representations:

\[ |+z\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad |-z\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \]

\[ |+x\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad |-x\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \]

\[ |+y\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \quad \text{and} \quad |-y\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}. \]

Choose all of the following statements that are true:

I. \( |+x|+x\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1 \)
II. \( |+x|+x\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \)
III. \( |+z\rangle(-x\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \)
IV. \( |+z\rangle(-x\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \)

A. I only B. I and III only C. I and IV only D. II and III only E. None of the above

CQS 3.1 A generic state is written in the \([+z], [-z]\) basis as \( |\chi\rangle = a|+z\rangle + b|-z\rangle \), but it can be written in another basis. In the \([+y], [-y]\) basis, \( |\chi\rangle = a'|+y\rangle + b'|-y\rangle \). Choose all of the following statements that are true:

I. \( a = (+z|x\rangle \) and \( b = (-z|x\rangle \)
II. \( a' \) and \( b' \) are the projections of \( |\chi\rangle \) along \( |+y\rangle \) and \( |-y\rangle \), respectively.
III. \( a' = (+y|x\rangle \) and \( b' = (-y|x\rangle \)

A. I only B. II only C. I and III only D. II and III only E. All of the above

CQS 3.2 A generic state is written in the \([+z], [-z]\) basis as \( |\chi\rangle = a|+z\rangle + b|-z\rangle \). Given that \( |+y\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \) and \( |-y\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix} \), choose all of the following statements that are true:

I. \( |+z\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad |-z\rangle = -\frac{i}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \)
II. \( |\chi\rangle = a \begin{pmatrix} 1 \\ i \end{pmatrix} + b \begin{pmatrix} 1 \\ -i \end{pmatrix} \)

A. I only B. II only C. III only D. All of the above E. None of the above

CQS 3.3 One way to accomplish the process of writing a state in a basis is by acting upon it with the operator \( \hat{I} \) written in that basis. Choose all of the following that are valid statements:

I. \( \hat{I} = |+x\rangle(+x\rangle \)
II. \( \hat{I} = |+x\rangle(+x\rangle + |-x\rangle(-x\rangle \)
III. \( \hat{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \) in any basis chosen

A. I only B. II only C. III only D. I and II only E. I and III only

CQS 3.4 Given that the identity operator \( \hat{I} \) multiplied by any state returns that state, choose all of the following that are equivalent way of writing the state \( |\chi\rangle \):

I. \( |\chi\rangle = \hat{I} |\chi\rangle = (|+x\rangle(+x\rangle + |-x\rangle(-x\rangle) |\chi\rangle \)
II. \( |\chi\rangle = \hat{I} |\chi\rangle = (|+y\rangle(+y\rangle + |-y\rangle(-y\rangle) |\chi\rangle \)
III. \( |\chi\rangle = \hat{I} |\chi\rangle = (|+x\rangle(+x\rangle + |-y\rangle(-y\rangle) |\chi\rangle \)

A. I only B. II only C. III only D. I and II only E. All of the above

CQS 3.5 If we want to express our state in the \([+x], [-x]\) basis, we can write \( |\chi\rangle = \hat{I} |\chi\rangle = (|+x\rangle(+x\rangle + |-x\rangle(-x\rangle) |\chi\rangle \). Choose all of the following that correctly represent the state \( |\chi\rangle = a|+z\rangle + b|-z\rangle \) in the \([+z], [-z]\) basis:

I. \( |\chi\rangle = (|+x\rangle(+x\rangle + |-x\rangle(-x\rangle) (a|+z\rangle + b|-z\rangle) \)
II. \( |\chi\rangle = |+x\rangle(a|+z\rangle + |-x\rangle(-b|-z\rangle \)
III. \( |\chi\rangle = a|+z\rangle + b|-z\rangle \)

A. I only B. II only C. III only D. All of the above E. None of the above

The final version of the CQS was implemented once online and twice in person. The CQS during the online implementation was presented as a Zoom poll with questions displayed via the “Share Screen” function, and for the in-person implementation, the poll was replaced by a functionally similar classroom clicker system. For each question, after displaying the polling results, the instructor held a full class discussion of the possible options provided. The Peer Instruction component was present in the in-person administration, but not the online administration because of difficulties in fostering small-group student discussion in the online environment. We note also that the instructors were different for the online and in-person classes.

To determine the effectiveness of the CQS, we developed and validated a pre- and post-test containing questions on topics covered in the CQS. The post-test contained small
changes in numerical values and the basis in which some states were expressed, but was conceptually similar. In both online and in-person classes, students completed the pre-test immediately following traditional lecture-based instruction on the topic. After administration of the CQS, students completed the post-test. Two researchers graded the pre-test and post-test, and converged after discussion on a rubric with inter-rater reliability greater than 95%. Selected pre- and post-test questions are reproduced below; some other questions are not examined here due to limited space. Question Q1 asks students to calculate the outer product given two ket states, which requires finding the corresponding bra state for one of them. Questions Q2 and Q3 ask students to express a given state in a different basis.

Q1. Given that in the \( z \)-basis
\[
|\chi\rangle = \frac{1}{\sqrt{3}} \left( \frac{3i}{5} \right) \text{ and } |\psi\rangle = \frac{1}{\sqrt{17}} \left( -4i \right)
\]
Find \(|\psi\rangle|\chi\rangle\). Show or explain your work to get credit.

Q2. Write the state \(|\chi\rangle = \frac{3}{\sqrt{10}} |+x\rangle + \frac{1}{\sqrt{10}} |-x\rangle\) in the form \(|\chi\rangle = a'|+z\rangle + b'|-z\rangle\) (that is, find \(a'\) and \(b'\)). Show or explain your work.

Q3. Write the state \(|\chi\rangle = \frac{3}{\sqrt{10}} |+x\rangle + \frac{1}{\sqrt{10}} |-x\rangle\) in the form \(|\chi\rangle = a'|+z\rangle + b'|-z\rangle\) (that is, find \(a'\) and \(b'\)). Show or explain your work. Note that we start here in the \( z \)-basis and wish to convert to \( x \)-basis, the opposite of the preceding question.

III. IMPLEMENTATION RESULTS

Though this study is quasi-experimental in design [37], prior work has shown that, compared to control groups given traditional instruction and homework, students had significantly better post-test performance when engaging with research-based tools [29,38]. The pre-test scores, post-test scores and effect sizes [39] are listed in Tables I-III. Table I gives results for the online class (\( N = 29 \)), and Tables II-III do so for in-person classes 1 (\( N = 25 \)) and 2 (\( N = 27 \)). Overall, the results are encouraging, and students performed well on the post-test, with reasonably impressive effect sizes. Below, we discuss some difficulties that were addressed after the administration of the CQS for all three years.

A. Outer product

Question Q1 asks students to calculate the outer product of two given states. Students’ most common mistake was to provide a scalar rather than a matrix, for which zero credit was given. This was observed for many students’ pre-test responses in all classes, and some students’ post-test responses during the online class. Students in the in-person years performed better on the post-test, with some even writing their answers in Dirac notation, indicating that these students were comfortable translating between the Dirac notation and matrix representation. The majority of students preferred using matrix representation, likely because the question had the given states in matrix representation and because it offers more compact notation. Some students neglected to take the complex conjugate when finding the corresponding bra state. Additionally, some students found the bra state corresponding to the ket state other than the one indicated, but they were given full credit if they otherwise performed the outer product correctly. Among incorrect responses, it was also common for students to provide the transpose of the correct matrix. On a related note, some provided the correct answer but showed for their work a nonstandard method for matrix multiplication, placing the row matrix to the left of the column matrix (which conventionally yields a scalar), indicating that they knew the result of such an operation but had not mastered the mechanics of matrix multiplication. It is possible that those students who unintentionally transposed the matrix also did not have a full understanding of the rules of matrix multiplication. All of these difficulties were much more prevalent on the pre-test than the post-test for all years.

B. Changing from \( x \)-basis to \( z \)-basis

Question Q2 provided a state in the \( x \)-basis and asked students to change to the \( z \)-basis. This question was deliberately left open-ended so that students could use the method with which they were most comfortable, as the CQS covered several distinct approaches to changing basis. Most students chose to directly substitute \([\pm x]\) states with their expressions in the \( z \)-basis, \(\frac{1}{\sqrt{2}}(\langle +z \rangle \pm \langle -z \rangle)\), which were given in the pre-test and post-test.

The most common mistake on the pre-test, other than a few students who left the question blank or made little progress, was to simply divide the expansion coefficients in the starting basis by \(\sqrt{2}\). Students with this type of difficulty did not recognize that such a state is not normalized. While the details of how students arrived at this result varied between students, many cases it may stem from discarding some of the inner products \((+x)|+z\rangle, (+x)|-z\rangle, (-x)|+z\rangle, \text{ and } (-x)|-z\rangle\) in their attempts to obtain the final answer, resulting in an incomplete projection along the new basis. Other students arrived at the conclusion that the expansion coefficients do not change when transforming from one basis to the other (i.e., \(|\chi\rangle = a'|+x\rangle + b'|-x\rangle = a'|+z\rangle + b'|-z\rangle\). These notions were largely corrected on the post-test.

Additionally, on the post-test, rather than finding \(a'|+x\rangle + b'|-x\rangle\) in the \( z \)-basis as asked, one student was observed to instead find the state \(b'|+x\rangle - a'|-x\rangle\), which is orthonormal to the given state. This was an interesting response, as it demonstrated an understanding of what makes an orthonormal basis, and possibly the idea that one can arbitrarily construct an infinite number of bases, though it had little to do with transforming from the given basis to the target basis.
Despite this, it is worse $(\mathbf{b} - \mathbf{a})$. Terms of taking projections along the new basis, which could of the identity operator or explicitly described their work in method, and students arrived at a correct answer on the post test, more difficult than question all of the information, so question students and cause cognitive overload as they try to process extra. Students who engaged with the CQS (N = 25).

<table>
<thead>
<tr>
<th>Q#</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>47%</td>
<td>60%</td>
<td>0.31</td>
</tr>
<tr>
<td>Q2</td>
<td>69%</td>
<td>84%</td>
<td>0.37</td>
</tr>
<tr>
<td>Q3</td>
<td>50%</td>
<td>86%</td>
<td>0.86</td>
</tr>
</tbody>
</table>

TABLE II. Results of the first in-person administration of the CQS (in-person class 1). Comparison of pre- and post-test scores, along with effect size as measured by Cohen’s $d$, for students who engaged with the CQS (N = 25).

<table>
<thead>
<tr>
<th>Q#</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>66%</td>
<td>94%</td>
<td>0.91</td>
</tr>
<tr>
<td>Q2</td>
<td>92%</td>
<td>98%</td>
<td>0.29</td>
</tr>
<tr>
<td>Q3</td>
<td>62%</td>
<td>98%</td>
<td>1.08</td>
</tr>
</tbody>
</table>

TABLE III. Results of the second in-person administration of the CQS (in-person class 2). Comparison of pre- and post-test scores, along with effect size as measured by Cohen’s $d$, for students who engaged with the CQS (N = 27).

<table>
<thead>
<tr>
<th>Q#</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>59%</td>
<td>85%</td>
<td>0.69</td>
</tr>
<tr>
<td>Q2</td>
<td>61%</td>
<td>87%</td>
<td>0.62</td>
</tr>
<tr>
<td>Q3</td>
<td>48%</td>
<td>78%</td>
<td>0.65</td>
</tr>
</tbody>
</table>

C. Changing from $z$-basis to $x$-basis

Like question Q2, question Q3 presented a state for students to change to another basis, this time starting in the $z$-basis and going to the $x$-basis. However, the relationships $|\pm z\rangle = \frac{1}{\sqrt{2}} \left( |+x\rangle \pm |-x\rangle \right)$ were not explicitly provided. Since it is possible to add and subtract the $|\pm x\rangle = \frac{1}{\sqrt{2}} \left( |+z\rangle \pm |-z\rangle \right)$ equations to find the explicit relationships, direct substitution is still a viable method. Some students went through this algebra, while others correctly recognized that the relationship between the $x$-basis and $z$-basis is symmetrical, which was another valid justification. These extra algebraic steps, however, can be challenging for some students and cause cognitive overload as they try to process all of the information, so question Q3 was considered to be more difficult than question Q2. Illustrating this, on the pre-test, many students (more than for Q2) were unable to make much progress with this question. In a similar vein, several students were observed to state that $|\psi\rangle = a|+x\rangle + b|-x\rangle = a|+z\rangle + b|-z\rangle$, with the expansion coefficients being the same in both bases, despite successfully working out the basis change for Q2. In all classes, a great majority of these students arrived at a correct answer on the post-test. Overall, for both questions Q2 and Q3, substitution was the dominant method, and fewer students used the spectral decomposition of the identity operator or explicitly described their work in terms of taking projections along the new basis, which could point to a lesser emphasis on these methods in class and on homework assignments.

IV. COMPARISONS BETWEEN ONLINE AND IN-PERSON IMPLEMENTATIONS

The pre-test scores were around 50-60% for all years, with some exceptions, most notably Q2 during in-person class 1. Similarly, with the exception of Q1 for the online class, the post-test scores were all quite high, at about 80-90%, with medium to large effect sizes for questions with greater improvements (see Tables I-III).

The biggest difference in this regard is observed in question Q1 during the online class, which is the only one that did not feature Peer Instruction. This question was on outer products, and Q1 scores for this class were lower than they were in the others. This could have been a concept that was more difficult for students to grasp, thanks to the difficulties posed by the online environment. Despite this, however, they went on to quite proficiently change the basis of two states in Q2 and Q3 after the CQS (see Table II). It is possible that the CQS was more effective at helping students learn change of basis than outer products in an online format, or that higher-quality in-person discussions helped students better understand outer products.

Online instruction results seem to be only slightly worse or no worse than in-person results. This is consistent with what we have seen in other studies [35,36]. One possible reason the online results without peer instruction are not significantly worse than the in-person results with peer instruction is that students had their cameras off; as such, some may have consulted resources even though they were not supposed to consult any [40].

V. SUMMARY

Validated clicker question sequences can be effective tools when implemented alongside classroom lectures. We developed, validated, and found encouraging results from implementation of a CQS on the topic of change of basis for two-state quantum systems. We find that students tend to struggle with outer products after traditional lecture-based instruction, typically trying to find a scalar or getting lost in non-standard matrix multiplication schemes. When changing basis, common mistakes included simply dividing the state by $\sqrt{2}$ or answering that the expansion coefficients are the same in both bases. These difficulties were much reduced on the post-test.

Following the administration of the CQS, post-test scores improved across both instructors and both modes of instruction, showing that the CQS can be helpful in a variety of situations. It is thus likely that the CQS can help reduce student difficulties on these concepts for any instructor who is interested in implementing it in their own courses.

ACKNOWLEDGEMENTS

We thank the NSF for award PHY-2309260.


Simulated Groupwork in an Asynchronous Course Learning about Radioactivity

Michael M. Hull
University of Alaska Fairbanks, College of Natural Science & Mathematics, Fairbanks, Alaska 99775, USA

Maximilian Jeidler and Eva Holzinger
Austrian Educational Competence Centre, Division of Physics, University of Vienna, Porzellangasse 4/2/2, Vienna, Vienna, Austria/1090

Despite its various benefits, one disadvantage that learners in asynchronous physics courses face is the lack of moment-by-moment responsive teaching that generally facilitates learning. Previous research has documented the use of simulated Peer Instruction in an asynchronous environment by presenting students on an individual basis with alternative responses authored by the instructor or collected from previous students. Inspired by this work, we created and implemented an asynchronous learning module utilizing simulated groupwork. In this module, individual students listen to recorded discussions between artificial group members who frequently pause to ask the student for input. The student then types a response before the module proceeds to the next prompt. Through a survey administered pre and post-instruction, we compare student learning with learning gains from in-person instruction on the same topic. Furthermore, using alluvial diagrams, we identify trends in pretest/post-test improvement and responses to particular module prompts to gain insight into how to improve the curricular materials. Although this is only a pilot study (\(N = 21\) from a class of 29 students in College Physics 2) with specific learning targets (1. that radioactive material is ubiquitous and 2. that irradiation generally does not cause something to become more radioactive than it already is), our findings support our hypothesis that simulated groupwork can be an effective means to promote student learning in asynchronous physics courses in general.
I. INTRODUCTION

One of the benefits of interactive engagement teaching methods, like Peer Instruction (PI) [1,2], is that they provide instructors with information on what their students are thinking, so that the instructor can teach responsively. Most interactive engagement practices operate under the premise that students will be gathered together during class to share their ideas with each other and with the instructor. Although this is usually a good assumption, it is not the case in asynchronous courses, where students access materials, including recorded lectures, at whatever timing is conducive to their schedules. Asynchronous learning is empowering to diverse populations of students who would not otherwise be able to take university courses; however, it is more difficult to have learning be responsive to student ideas when it is asynchronous (e.g., [3]). In recognition of this, Englund et al. recently created simulated PI [4]. In simulated PI, after an individual student provides a personal answer to a ConcepTest, answers of simulated classmates arguing for the alternative selections are presented to the student. The student then has the opportunity to change his or her mind before the correct answer is revealed. This is at least semi-responsive in the sense that the alternative reasoning is presented after the student has submitted an initial answer, and it presents the student with an opportunity to revisit the initial answer (see also [5]).

Our study is similar to that of Englund et al., except that our investigation concerns simulated groupwork that could potentially substitute group-based recitations in asynchronous physics classes. We created our simulated groupwork to enhance student understanding of radioactivity as part of College Physics 2 at the University of Alaska Fairbanks in spring 2023. Our research questions are 1) how successful was the simulated groupwork? and 2) how might the simulated groupwork be improved?

1.1 Student understanding of radioactivity: “get it away from me!”

Although education researchers have documented numerous student ideas about the topic of radioactivity (see [6] for a comprehensive literature review or [7] for an abridged review, including [8–15]), the “undifferentiated view” has received the most attention. Learners who are “differentiated” correctly understand that radioactive substances are distinct from the radiation they send out. While the radiation can cause damage to a victim by removing electrons, interactions of the radiation with the nuclei of victim atoms is sufficiently rare to generally be negligible; as such, irradiation does not generally make other objects radioactive. The “undifferentiated” view, on the other hand, lumps radioactive material and radiation—and consequently, irradiation and contamination—together into one dirt-like or even germ-like substance referred to by some education researchers as “it”. Research has also shown that students tend to view “it” as, generally speaking, being dangerous and man-made.

II. METHODOLOGY

The simulated groupwork we created was based upon Johnson’s Inquiry into Radioactivity (IiR) curriculum [16], which demonstrated success at developing the differentiated view in non-science majors at Black Hills State University [14]. Although IiR was originally designed to fill a physical sciences elective for non-science majors, MS.Ed. students taking Hull’s seminar at the University of Vienna condensed the first part of IiR into a version that would occupy only three class periods [17]. In these three lessons, students are told that a Geiger Mueller (GM) counter is a device that they have seen on TV nearby Fukushima measuring radiation from radioactive sources. Learners explore the world around them with the GM counter and are generally surprised to learn (ideally) that radioactivity is ubiquitous (although some things are much more radioactive than others), and the term “background radiation” is introduced. Once students establish a baseline for how much the GM counter clicks in an empty classroom (the range of values for minute-long readings), they are then ready to put a strawberry into close proximity of a strong radioactive source to see that, even after 48 hours of being close to the uranium glass, the strawberry does not register an increased number of clicks on the GM counter. In contrast to typical student predictions, students recognize (ideally) that the radiation did not make the strawberry radioactive. For use in online classrooms or in classrooms that do not have GM counters, we created videos of these experiments.

II.1 Implementation and assessment of Inquiry into Radioactivity in College Physics 2

For the asynchronous College Physics 2 class, Hull created and audio recorded a dialogue between two virtual students, Alex and Bailey, as they progressed through the three condensed IiR worksheets. This dialogue was based upon personal experience of listening to typical in-person student responses to the prompts on the worksheets and knowledge of common student difficulties documented in research literature. Hull played this audio file while screen recording the various videos and scrolling through the IiR worksheets. For each prompt in the worksheets, Alex and Bailey ask the student for his or her own opinion. Using PlayPosit, we had the module stop at each of these questions and only continue once students submitted a response. This activity took the place of weekly online lectures and homework and students were awarded points for participation. Almost all students in the class completed the module (28 students out of 29).

Improvement in student understanding was assessed using a survey developed by Holzinger as part of her MS thesis [18]. The survey includes three items to assess student
III.2 Investigating room for improvement

To address our second research question, RQ2) how might the simulated groupwork be improved?, our focal point was the recognition that, as discussed in the introduction, asynchronous instruction generally is lacking in responsiveness to student ideas. It is possible to alleviate this somewhat by embedding branching points into PlayPosit modules. In the simplest version of such a branching point, students answer a multiple choice question. If they answer incorrectly, they are brought to additional instruction that is skipped by students who answered the prompt correctly. This is somewhat analogous to instructors checking in with students during groupwork and giving additional guidance if needed. Such a systematic method is less than ideal, as students are quick to realize if instructors are only checking in when they are incorrect; nevertheless, most instructors are most likely to intervene at such times, and so we approach our module improvement with this method as a first step.

With the desire to find the most efficient locations to introduce branching points, Hull developed and carried out a five-step process: 1) comb through student responses to the module prompts one by one and flag each response where an instructor should intervene, were the lesson to take place in a synchronous classroom; 2) identify the ten prompts (or, in case of a tie, the smallest number of prompts more than ten) in the module that received the most flags; 3) for each of these most-problematic prompts, identify the learning goal (for example, reaching the differentiated view) that the prompt most closely corresponds to; 4) identify all (e.g.) DIFF items on the survey and choose a criterion by which growth will be determined (here, due to small sample size, we coded a student as “differentiated” if all three DIFF items were answered correctly); 5) create alluvial diagrams (we used R Studio) to demonstrate student progression from the pretest (coded, e.g., as “differentiated” or “undifferentiated”) to the module prompt (coded for whether the instructor should “intervene” or “pass” by the group discussion) to the post-test (coded the same way as the pretest). We will elaborate upon alluvial diagrams and this last step.

Alluvial diagrams simplify and highlight structural changes between clusters so that transition rates can readily be seen. Alluvial diagrams are a type of Sankey diagram, which are used in a variety of research fields, including physics education research [22–24]). In our study, there are eight possible alluvia for each [pretest – module prompt – post-test] combination. Our names for alluvia discussed in the following paragraphs are labeled in Figures 1 and 2 in the Results section.

The substeps in step 5 for module prompts pertaining to the differentiated view are as follows: 5i) check if the “ceiling” alluvium [differentiated – pass – differentiated] dominates over all eight alluvia, indicating that the prompt is perhaps redundant and a candidate for removal to streamline the module; 5ii) from the four alluvia beginning with a “differentiated” pretest code, check the size of the “misled” alluvium [differentiated – pass – undifferentiated]. If this alluvium is larger than the “ceiling” alluvium, then the prompt may be misleading students; 5iii) compare the four alluvia beginning with an “undifferentiated” pretest code. If the learning goal is particularly ambitious, the “baffled” alluvium [undifferentiated – intervene – undifferentiated] will dominate these four. If, on the other hand, great gains were seen in the survey pretest/post-test, then, ideally, the dominating alluvium will be the “victory” alluvium [undifferentiated – pass – differentiated]. If this alluvium is the largest, or if it is the second largest with the “baffled” alluvium in the lead, then the prompt is most likely helpful with learning. If “victory” is the largest but “baffled” is also large, then the module prompt is one that can be potentially
be improved with a branching point. If, instead, one of the two “other” alluvia is the largest of these four, or if “baffled” is the largest followed by an “other” alluvium, then the prompt is most likely not directly tied to the supposedly associated survey items and it should perhaps be replaced.

III. RESULTS

In the post-test survey, we asked students for their impressions of the learning module. The majority of students reported either neutral or positive feelings regarding the module overall (19/21) and about the simulated groupwork being better than pre-recorded lectures (15/21). Feedback included “…Breaking the material down piece by piece in the form of a conversation made it much easier to grasp”. Our first research question was RQ1) how successful was the simulated groupwork? In answering this question, we did not assume that the three DIFF items are equally difficult, and so we did not calculate an average score across the items (likewise for the UBIQ items). In future large-N studies, we can perform Rasch analysis to determine the relative difficulty of each item for a more nuanced analysis.

For our pilot study, we applied the code, e.g., “differentiated”, if all three DIFF items were answered correctly; otherwise, we applied the code “undifferentiated”. First, regarding the differentiated view, we saw improvement comparable to that of IIR when used as a full-semester in-person class [14]. Two of the 21 students left one or more of the DIFF items blank. From the remaining students, we saw a growth of 3/19 to 16/19 differentiated students post-instruction [21]. Regarding understanding that radioactivity is ubiquitous, we saw much more modest (not statistically significant) growth, from 4/20 to 7/20 students answering all three items correctly (one student left one of the three items blank). The hardest of the three items by far was UBIQ2. Whereas our growth was 4/20 to 11/20 correct respondents, the MS thesis of Jeidler found that students answering this item correctly also remained below 10% post-instruction. As this is only a pilot study and our sample size is small, we can only tentatively answer our first research question with the following: regarding the differentiated view, our lesson utilizing simulated groupwork was comparably effective to the original IIR which features genuine groupwork. Regarding the idea of radioactivity being ubiquitous and not necessarily artificial and dangerous, our lesson was comparably effective to other in-person instructional approaches covering radioactivity. However, in light of the emphasis the module places on using the GM counter to measure the radiation coming from everyday objects, we are optimistic that greater learning gains can be had from improved versions of the learning module. This motivates our second research question, RQ2) how might the simulated groupwork be improved?

To answer this second research question, the Hull examined the responses of the 21 students on the module, prompt by prompt. In total, there are 45 prompts in the module. Hull flagged one prompt, FUKUSHIMA, as being the most problematic, as 10 of the 21 students responded in a way that warranted instructor intervention (all other prompts had fewer of these flags). FUKUSHIMA asks students, after they have acknowledged that the strawberry is not more radioactive despite being in close proximity to uranium glass for 2 days, “What might be the cause for the increased radioactivity measurements near Fukushima?” A response that was coded as “pass by without intervening” was “… perhaps the radioactive material did not ‘make’ the ground radioactive, but just coated the ground in radioactive material…” On the other hand, a response coded as “intervention needed” was “the radioactive material was much stronger at Fukushima.” We considered that this might be a good location for a branching point; namely, to replace the currently free-response FUKUSHIMA prompt with a multiple-choice version. Students who select that the increased GM counter readings were due to dust coating the ground near Fukushima would pass on with the lesson.

Students who selected that the cause is higher radiation levels would listen to additional dialogue between Alex and Bailey, where the idea of breaking the uranium glass and embedding parts of it into the strawberry would be mentioned. Since this would require additional time for the learner, we first examined the alluvial diagram (Figure 1) for [DIFFERENTIATED.PRE – FUKUSHIMA – DIFFERENTIATED.POST] to see if performance on this prompt correlated with growth on the DIFF items.

FIG. 1. Alluvial diagram for [DIFFERENTIATED.PRE – FUKUSHIMA – DIFFERENTIATED.POST]. The top boxes are desirable responses: (from left to right) “correct” on all three DIFF items on the pretest; “pass” on the FUKUSHIMA prompt; “correct” on all three DIFF items on the post-test.
Here, each of the 19 respondents fell into one of five alluvia. The “victory” alluvium dominates, suggesting that this prompt might be important for student learning. Specifically, this alluvium shows students who, on the pretest, demonstrated the undifferentiated view by answering at least one of the three DIFF items incorrectly. The students “victoriously” answered FUKUSHIMA correctly and (at least plausibly) left the module with an improved understanding, as demonstrated on the post-test items. However, there is also a noticeable (three students out of 19) “baffled” alluvium, indicating that this, indeed, would be an appropriate prompt to build in a branching point. Had these three students received additional support on FUKUSHIMA, it is plausible to think that they too would have joined the “victory” alluvium.

Seven of the 12 most-problematic prompts were, like FUKUSHIMA, relevant to the differentiated view. Three of these seven prompts had alluvial diagrams similar to Figure 1 and were thus found to be candidates for branching points. The other four module prompts related to the differentiated view had no students in the “baffled” alluvium and so were judged to be sufficiently effective as is. Regarding seeing radioactivity as ubiquitous, the majority of the relevant module prompts were relatively issue-free for students. Only one of the 12 most-problematic prompts, INFRARED, was potentially related to this learning goal. INFRARED asks students to watch a video of a person rub his hands together and hold them over the GM counter and in front of an IR camera. Although this prompt is originally intended to help strengthen the differentiated view by showing that there are types of radiation that are unrelated to radioactivity, it should, ideally, also show that people, like other everyday objects measured, make the GM counter click. An example of a response coded as sufficient to allow the teacher to “pass” by without intervening was: “The heat detector measured the heat in the person’s body, but there were not more clicks with increased heat.” On the other hand, a response coded as requiring a teacher to “intervene” was “It seemed like the less heat around the detector, the less clicks that could be heard.” This latter statement could reflect the idea that clicks decrease when the person comes near since people are not radioactive. To see whether this prompt affected student learning, we created an alluvial diagram (Figure 2).

Here, the “victory” alluvium does not dominate, nor does the “baffled” alluvium. Rather, of the four alluvia beginning with an incorrect response on the pretest, an “other” (“other3”) alluvium is the widest. This suggests that the prompt is not closely connected to the learning goal. Since none of the other 12 most-problematic module prompts were related to the ubiquitousness of radioactivity, it was decided that other, more explicit prompts should be introduced to the learning module. For example, when students are watching videos of everyday objects being measured by the GM counter, a video of a person being measured (without the infrared camera) should also be added. The remaining 4 most-problematic prompts were related to the idea that it is random when an individual nucleus decays, the focus of our prior work [7,25–28].

![Alluvial diagram](image)

**FIG. 2. Alluvial diagram for [UBIQUITOUS.PRE – INFRARED – UBIQUITOUS.POST].** The top boxes are desirable responses: (from left to right) “correct” on all three pretest UBIQ items; “pass” on INFRARED; “correct” on the three post-test UBIQ items.

### V. DISCUSSION AND CONCLUSION

To the best of our knowledge, we are the first to attempt to make instruction for asynchronous learners more responsive by simulating groupwork. Although this was only a pilot study, we found that most students successfully came to understand that radioactive objects do not, generally, make other objects radioactive in the way that germs or dirt spreads. Furthermore, we found improvement in students realizing that everything is already a little bit radioactive. At the same time, we identified room for improvement of the module by looking at patterns of student responses on survey items and corresponding module prompts.

We find the results of our pilot study encouraging, as we obtained conceptual gains comparable to in-person instruction on radioactivity [14,21]. Our work was specific to student learning about radioactivity, and future work should investigate the creation and assessment of simulated groupwork for other physics topics. The procedure discussed in the methodology of visually inspecting alluvial diagrams is a generic one that can potentially be used as a starting point for improving curriculum in any topic. With a greater number of respondents, qualitative content analysis can be used to categorize student responses beyond the binary “instructor intervention needed” or “instructor can pass by without intervening”. Similarly, finer-grained coding can be applied to the survey items. This analysis is important for optimizing asynchronous curricular content to be responsive to learners without being unnecessarily burdensome to them.
[18] E. S. Holzinger, The Public’s Knowledge on Radioactivity in Austria Affected by the Time Passed After Graduation and the Type of School Attended. The Public’s Knowledge on Radioactivity in Austria Affected by the Time Passed After Graduation and the Type of School Attended, MS Thesis, University of Vienna, 2022.
Drawing on force ideas for kinematic reasoning in introductory physics

Tra Huynh (she/her)
Department of Physics and Astronomy, Western Washington University, 516 High Street, Bellingham, WA 98225

Anne Alesandrini (they/them), Lauren C. Bauman (she/her), and Olin Sorby (he/him)
Department of Physics, University of Washington, Box 351560, Seattle, WA 98195

Amy D. Robertson (she/her)
Department of Physics, Seattle Pacific University, 3307 Third Ave W, Seattle WA 98119

In this paper, we identify some of the connections students make between force ideas and kinematics concepts, in their responses to kinematics questions. We coded 887 written responses to three different kinematics questions and identified patterns where students draw on force to make sense of kinematics concepts such as acceleration, velocity, and trajectory. We found that students draw on force frequently, in addition to other kinematics reasoning, in resourceful and context-dependent ways. Our findings suggest that instruction may be able to productively make use of students’ understanding of force to support kinematics learning.
I. INTRODUCTION

Prior work has studied student understanding of forces and kinematics, showing that the two concepts are strongly related and support each other in university introductory physics contexts. Extensive research has identified patterns in student understanding of force and kinematics through various lenses, including naive beliefs and misconceptions [1-2], difficulties [3-5], knowledge in pieces [6, 7], and conceptual resources [8, 9]. Regardless of the approach, research tends to agree that students come to physics class with prior sensory experiences of force and motion, although their conceptualization of force might not be aligned with current physics models. However, instructors’ stances towards students’ understanding of force and motion—i.e. whether students’ ideas are robust and resistant to instruction [10] or they are potentially productive and context-dependent [8,9]—inevitably influence instructional perceptions and design [11, 12] around the two concepts in physics courses.

Traditional introductory textbooks and curricula often teach kinematics before forces (see [13-15], for example). In this approach, acceleration is taught from a kinematical perspective before introducing its mathematical relationship with force. Then, when force is introduced via Newton’s 2nd Law, force is presented mathematically as something that produces acceleration. There are fewer approaches where force is introduced qualitatively or even quantitatively before the presentation of kinematics (see [16], for example). These latter approaches introduce forces first, and then introduce acceleration as the result of an imbalance of forces. An instructor’s choice of approach might depend on their instructional goals for student learning. For example, when instruction focuses on solving kinematics questions using definitions and graphical representations, instruction might postpone—and even discourage—the use of force ideas. There is little research on the relative effectiveness of a kinematics-first vs. a forces-first approach to instruction, with the exception of [17], which found that there were no differences in conceptual and attitudinal survey outcomes between the two approaches.

Our study builds on [17], contributing to this body of work from a slightly different angle. We take a resources approach [6,7] to identifying ways in which students draw on force reasoning to productively sense-make about kinematics questions. We analyzed student responses to kinematics questions that did not specifically call for force reasoning to (i) get a sense of how frequently students use force ideas in kinematics contexts and to (ii) characterize some of the ways in which students draw on force reasoning to support kinematics reasoning. We found that even without prompting, students often draw on ideas about forces in their responses as they think about acceleration, velocity, and trajectory, in addition to using other kinematics information including graphs, diagrams, vectors, and mathematics equations. Our findings suggest that force ideas can be conducive for student understanding of kinematics, implying that it is worth considering instruction that does not discourage force reasoning in kinematic instruction.

II. THEORETICAL FRAMEWORK: RESOURCES

In resources theory, a resource is a piece of knowledge that is activated in context-sensitive ways, sometimes in concert with other resources, to form an idea, explanation, argument, or theory [7, 12, 18-24]. Researchers have theorized extensively about the development, structure, and role of resources, and have used resources theory to highlight the dynamic, emergent, complex-systems-like nature of student thinking.

Our work draws extensively from resource theory’s orientation toward student thinking as fundamentally sensible and continuous with formal physics [6, 12, 18-20, 22, 24], seeking to make apparent the continuities between students’ thinking and formal physics, even and especially when that thinking does not use the language of formal physics or is canonically incorrect. Our work also takes up resource theory’s definition of learning, which involves changing the structure or activation of resources, by reorganizing, refining, or increasing the degree of formality of resources [6, 20-23]. Our primary aim in identifying resources is to provide instructors with knowledge that they can use to build from student ideas in instruction.

III. RESEARCH CONTEXT AND METHODS

A. Context

The resources we report here were identified in written student responses to three kinematics conceptual questions: the Ball-on-ramp, the Oval track, and the Comet questions. All three questions included diagrams (shown in Fig. 1) and asked for student reasoning about the objects’ velocity and acceleration at specific points along the trajectories. For example, the Oval track question showed students the top-view diagram of a car moving at constant speed along an oval track and asked students to draw vectors to represent the velocity and acceleration of the car at points A through F and explain their reasoning.

We chose these questions because they are questions or modified versions of questions that have previously been used in studies that investigate students’ kinematics ideas from different theoretical lenses. Student can use kinematics rules to infer the object’s velocity and acceleration in each of these scenarios, for example by finding the displacement or by subtracting velocity vectors, respectively. However, we found that student responses frequently included connections to force to make sense of the kinematics concepts, even though unprompted by the questions.
We analyzed a total of 887 written responses from introductory physics courses at four US colleges and universities. Three of the four institutions are in the Pacific Northwest; one of them is a large public university, one is a small private university, and the other is a mid-size community college. The fourth institution is a large public university on the East coast. The racial and/or ethnic demographics for the colleges/universities in our study versus all college/university students are shown in Figure 2. Figure 2 suggests that the institutions in our study are not racially and/or ethnically representative of the population of college-bound freshmen in the US. By this measure, the universities in our study serve more Asian and Asian American students, fewer Hispanic or Latinx students, fewer Black or African American students, fewer multiracial students, and fewer white students than the general population of college students. The median parental income of the students at colleges/universities in our study is also higher than the national average. This sampling limits the generalizability of our results; though the resources we identified are common among the students in our sample, we cannot speak to their commonality in the population of introductory physics students writ large.

B. Method

To identify how students in our sample used force ideas to reason about kinematics concepts, we conducted a preliminary coding of student responses to the three questions. First, we made note of every statement that included a force idea and a kinematics idea. Then we built a coding scheme [27] to consistently identify the kinematics concepts that students name in relation to force across the three questions. We found that among all kinematics concepts, the three concepts that were most frequently connected to force were acceleration, velocity, and trajectory of motion. Using the coding scheme, authors TH and AA then independently coded 20% of students’ responses, first identifying whether a response included a force idea (yes/no). If the response included a force idea, we coded for the kinematics concepts that students connected force ideas to, if any, including acceleration, velocity, and trajectory. We measured inter-rater agreement using the normalized difference between the total number of possible codes and total number of disagreements because our codes are not mutually exclusive [28, 29]. The percentage agreement was 98.2%. Disagreements were resolved through discussion, resulting in a modified coding scheme which author TH used to code the rest of the data set. Lastly, author TH used each set of responses coded with force ideas (forces connected to acceleration, forces connected to velocity, and forces connected to trajectory) to identify emergent patterns in the ways that forces ideas were used resourcefully to reason about these kinematics concepts.

IV. RESULTS AND DISCUSSION

A. Force ideas are prevalent in kinematics sense-making

We found that students frequently used force ideas to reason about the three kinematics questions in our data set (Table I). Oftentimes, students linked force ideas to more than one kinematics concepts, resulting in the assignment of multiple codes to a single response. Other contextual factors, including class-size, instructional styles, formal and informal exposure to force concepts, and lived experiences with force ideas, may impact the prevalence and particularities of students’ use force ideas; however, we do not have sufficient information to formulate specific claims about the influence of specific contextual factors.

Though not the focus of this paper, we found that students often drew on force reasoning in concert with reasoning with kinematics definitions and representations, including graphs, vector diagrams, etc. For example, a response to the Ball-on-ramp question stated:

"When the ball is rolling up the ramp, it is slowing down. This is because the direction of the acceleration is opposite to the direction of motion of the ball. Because of this, the velocity vectors are getting smaller. [...] The acceleration..."
has a constant magnitude throughout because the acceleration acting on the ball is the acceleration due to gravity. This is 9.8 m/s^2 and does not change.

In this response, a student used the relative direction of acceleration and motion (“opposite”) to explain why “velocity vectors are getting smaller,” demonstrating conceptual understanding of multiple kinematics concepts, including the vector nature of velocity and acceleration. Then they specifically referred to gravity as the reason for the constant value of acceleration (underlined). This is an example of how drawing a force idea can complement—rather than hinder or replace—student kinematics reasoning.

### TABLE I. Prevalence of force ideas in responses to kinematic questions (Force, F-A, F-T, F-V are codes that identify a force idea and connect forces with acceleration, trajectory, and velocity respectively).

<table>
<thead>
<tr>
<th>Question</th>
<th>Ball-on-ramp</th>
<th>Oval</th>
<th>Comet</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>327 (100%)</td>
<td>271 (100%)</td>
<td>289 (100%)</td>
</tr>
<tr>
<td>Force</td>
<td>123 (37.6%)</td>
<td>29 (10.7%)</td>
<td>95 (32.9%)</td>
</tr>
<tr>
<td>F-A</td>
<td>91 (27.8%)</td>
<td>16 (5.9%)</td>
<td>71 (24.6%)</td>
</tr>
<tr>
<td>F-T</td>
<td>35 (10.7%)</td>
<td>19 (7.0%)</td>
<td>17 (5.9%)</td>
</tr>
<tr>
<td>F-V</td>
<td>45 (13.8%)</td>
<td>6 (2.2%)</td>
<td>32 (11.1%)</td>
</tr>
</tbody>
</table>

### B. Drawing on force ideas to reason about acceleration

Among all kinematics concepts in our data, students in our sample drew on forces more frequently when reasoning about acceleration than when reasoning about trajectories or about velocity (Table I). Sometimes they connected force ideas to reasoning about the magnitude of acceleration, other times to its direction.

#### 1. Force causes acceleration

The resource “force causes acceleration” is identified when students name a causal relationship between force and acceleration. Oftentimes, students’ responses include phrases such as “acceleration due to force” or “force that causes acceleration”, signaling this resource. For example, a response to the Ball-on-ramp question stated:

“[…] The direction [of acceleration] can be easily explained by gravity. Gravity is always down and since that is the only outside force on the ball it is the only thing causing acceleration.”

In this response, the student determines the direction of acceleration by drawing on the idea that “[gravity] is the only thing causing acceleration.” Here, it is in identifying the force first that the student correctly characterizes the acceleration as down; thus, in this case, associating acceleration with force is the means by which the student answers correctly.

#### 2. Acceleration is (part of) force

We found that sometimes students use acceleration as an equivalent term to force: either acceleration is force or acceleration is part of a force exerted on the system. For example, in response to the Oval track question, a student wrote:

“Acceleration towards the interior of the track is the friction force that allows the car to turn on the track. Where the curve [is] sharpest/tightest [Point B and Point A] the net force/acceleration allows the car to turn. At Point C, the curve is not as strong, requiring less force towards the interior to stay on the track.”

In this response, the student specifically refers to acceleration as though it is the force that allows the car to turn. Although the student incorrectly equates acceleration and force, they correctly identify the points on the oval trajectory where acceleration is largest/smallest, justifying their choice on the basis of the magnitude of the force. This is an example of how students might use the terms force and acceleration interchangeably in ways that at least do not hinder their learning, and may be productive or helpful.

### C. Drawing on force ideas to determine trajectories

Students also frequently draw on forces to reason about trajectories of motion, often arguing that forces cause objects to take certain trajectories.

#### 1. Gravity pulls objects down

Students often consider the impact of forces on the trajectories of objects, and more specifically, that gravity pulling objects results in certain trajectories. For example, responding to the Ball-on-ramp question, a student wrote:

“The main force acting on the ball at the instant the ball is pushed up is gravity. Gravity will pull down objects to the lowest point possible so the acceleration of gravity will point downwards (down the ramp) […].”

This is an example where a student draws on the idea of force (gravity) to make sense of the ball’s trajectory; specifically, gravity causes the ball to roll down to “the lowest point possible.” Although the student did not arrive at the canonically correct answer (perhaps due to their not considering the normal force), in drawing on the impact of force, the student was able to consistently predict the resulting trajectory, which was helpful for the student’s reasoning about acceleration. Although students most often drew on this resource in the context of the Ball-on-ramp question, connections between gravity and trajectory were also found to be productive in the context of the Comet question, in which students think the gravitational force by the sun is what pulls the comet towards the sun. Although the Comet question and the Oval track question share the feature of circular motion, gravity and trajectory is found to show up more often in the Comet question, likely as a result of the apparent mass at the center of the trajectory.

#### 2. Force is needed to keep an object on curved trajectory

Students often drew on the idea that forces are needed to keep an object on its trajectory. This resource is particularly
common in the Oval track question and the Comet question, where the trajectories are curved (e.g. oval and circular). For example, a response to the Oval track question stated:

“The magnitude of the acceleration also depends on the slope of the track since when it is ‘flatter’ there is not much force that needs to keep the car in circle while when the track is ‘steeper’ or more round there is a greater force required to keep the car in its trajectory.”

In this response, the student draws on the connection between force and trajectory to reason about the change in the magnitude of forces throughout the oval trajectory; specifically, force is greater when the object takes a “steeper” curve and less so when the object takes a “flatter” curve. In using this resource, students were able to make inferences about the magnitude of the acceleration at different points along the trajectory.

C. Drawing on force ideas to reason about velocity

Similar to previous studies [8, 9], we found that students often connect force and speed.

1. Force causes objects to speed up/slow down

Students commonly drew on the idea that force has an impact on how an object’s speed changes. Although we found students incorrectly relied on ideas such as a “larger force results in a larger speed,” drawing on force was resourceful in certain ways for the students in our sample. For example, in response to the Comet question, a student wrote:

“Since the [comet] is moving away, the gravitational force causes it to slow down to ‘turn around’ until it reaches A and speeds up again.”

In this response, the student makes sense of the comet’s motion by drawing on the gravitational force between the sun and the comet. Specifically, when the comet moves toward the far end of its trajectory from the star, the impact of the gravitation force now is thought to make the star slow down. In this example, although the student did not specify the change in the magnitude of the gravitational force and its impact on the comet’s motion, drawing on ideas of force informally is still productive for the students to correctly conceptualize the speed of the comet through the curviest part of its trajectory (slow down then speed up).

2. Going against (net) force slows down object

Students often drew on the resource that an object will slow down if its motion is opposite the direction of the net force. This idea is particularly common in the Ball-on-ramp question, which might be due to the force direction being constant (downward) in this case. For example, one student wrote:

“The higher up the ramp gets, the slower the ball will roll as there is gravity and possibly some friction acting upon it. As it goes down, the ball is no longer going against gravity, rather going with it which is why the ball is picking up speed as it goes down.”

In this case, the student correctly identifies the forces exerted on the ball (gravity and friction), however, miscounts for friction in the following step. Identifying a relevant force (gravity) supported the student in reasoning how the object changes its speed up and down the ramp.

V. CONCLUSION AND IMPLICATIONS

In kinematics instruction, instructors may have specific instructional goals for student learning, such as practice using graphical information and/or vector rules to find acceleration and velocity. In some cases, these learning goals prompt instructors to discourage students from using force reasoning in the context of kinematics, and to wait to introduce forces formally until after the kinematics unit. This study demonstrates that students in our sample frequently draw on force reasoning spontaneously in responses to kinematics questions, in many cases in ways that support them in correctly answering the question or forming generative connections between forces and kinematics concepts or forging relationships among kinematics concepts such as between acceleration and trajectory. Future studies might further investigate the impacts of contextual factors on students’ use of force ideas and the affordances of force reasoning in kinematics contexts, i.e. whether students who use force reasoning more often answer kinematics questions correctly than students who do not, or whether there are particular contextual factors that shape the helpfulness (versus hindrance) of force reasoning in kinematics contexts. Future work might also explore whether using force reasoning in kinematics concepts changes the landscape of student attention—e.g., do students who use force reasoning focus less on kinematics representations and more on changes in motion.

In general, we interpret our findings as an existence proof that using force reasoning in kinematics does not necessarily disadvantage students from understanding kinematics. In fact, force ideas often served as resources for reasoning about kinematics for the students in our sample. Our findings suggest that kinematics instruction might benefit from building upon students’ spontaneous ideas of force, rather than discouraging it.

ACKNOWLEDGMENTS

We are thankful to Lisa Goodhew, Rachel Scherr and the anonymous reviewers for feedback on the manuscript and to the instructors and researchers who helped with data collection. This material is based upon work supported by the National Science Foundation under Grant No. 1914572 and 1914603.


156
Toward helping students develop error detection skills

Safana Ismael  
*Department of Physics, North Dakota State University, 1340 Administration Ave, Fargo, North Dakota, 58105*

Mila Kryjevskaia  
*Department of Physics, North Dakota State University, 1340 Administration Ave, Fargo, North Dakota, 58105*

Recent findings suggest that even those students who demonstrate relevant formal knowledge tend not to use it productively, especially on tasks that elicit intuitively appealing incorrect responses. Dual-Process Theories of Reasoning suggest that to catch a mistake, reasoners must engage in the process of error detection and override: recognize reasoning red flags, consider alternatives, and apply relevant knowledge to check their validity. It is, however, challenging for many novice physics learners to recognize what specific formal knowledge must be used as a criterion that needs to be satisfied for validating or rejecting a response. To help students develop skills necessary for error detection and override, we designed a sequence of systematic spaced practices in the context of Newton’s 2nd law. We examined the effectiveness of this approach and identified specific factors that contribute to more productive engagement in error detection and override.
I. INTRODUCTION

Knowledge is central to productive reasoning but insufficient, especially in situations that present reasoning hazards [1–4]. Consider two students in an introductory mechanics course, Lisa and Danny (all names are pseudonyms). The students discuss forces acting on a box at rest on a horizontal surface while a constant horizontal 30N force is applied to the box, as shown in Fig. 1a [5,6]. Both students correctly argue that since the box is at rest, according to Newton’s 2nd law, the force of friction must be 30N to the left. However, Lisa and Danny abandoned this line of reasoning on the follow-up question in Fig. 1b, where two identical boxes are now at rest on surfaces with different friction coefficients ($\mu_A < \mu_B$) while a horizontal 30 N force acts on each box. Lisa now argues that the force of friction on Box A is less than that on Box B because $\mu_A < \mu_B$. Danny agrees and supports this response with the expression for the maximum value of the static friction, $f_r = \mu N$. While this expression is not applicable in this case, it provides confirmation for Lisa’s intuitively appealing but incorrect response. Neither student seems to question the validity of their answers by checking for consistency with Newton’s 2nd law that they had just applied on a nearly identical problem presented without salient distractive features (i.e., different $\mu$).

On question 1, Lisa and Danny demonstrated the knowledge and skills necessary to analyze forces acting on an object at rest. However, they did not transfer this knowledge to solve question 2 correctly. Inconsistent responses like Lisa’s and Danny’s often persist even after instruction [3,7–11]. To help students minimize reasoning inconsistencies, it is necessary to 1) understand the cognitive mechanisms responsible for productive and unproductive reasoning pathways and 2) pinpoint factors and instructional circumstances that help students enhance their reasoning skills necessary to validate or reject a response.

In this paper, we use Dual-Process Theories of Reasoning (DPToR) as a theoretical framework [1,2,12]. We describe a sequence of instructional interventions informed by DPToR and examine the roles of two factors that may impact student reasoning: the strength of relevant knowledge and the tendency toward cognitive reflection.

II. THEORETICAL FRAMEWORK

Research in cognitive psychology suggests that reasoning involves two processes: quick and subconscious process 1 and slow and deliberate process 2. Process 1 is often referred to as "gut feeling" or intuition. It immediately recognizes a given situation in a specific way based on prior knowledge, experiences, and expectations. We agree with the parsimonious and pragmatic definition of intuition as “nothing more and nothing less than recognition.” This recognition (often much more accurate for experts than novices) leads to the provisional mental model, which becomes available for scrutiny by the slow, deliberate, and analytical process 2, as shown in Fig. 2. It could be argued that Lisa and Danny immediately and accurately recognized the task in Fig. 1a as “about balanced forces.” However, the salience of different $\mu$ in Fig 1b may have overshadowed this approach and cued a provisional model involving static friction based on $\mu$.

Process 1 cannot be turned off. As such, a provisional mental model is an entry point into any reasoning path, and process 2 is tasked with evaluating its validity. If a reasoner is confident in the provisional model, process 2 may be entirely circumvented such that a conclusion is reached via a path of cognitive frugality. Knowing when it is safe to jump to a conclusion via that path is linked to cognitive reflection skills, defined as a reasoner’s tendency to mediate incorrect intuitive responses by engaging in process 2 analysis. The Cognitive Reflection Test (CRT) is often used to measure domain-general cognitive reflection skills [13–16].

If Process 2 intervenes, it still may not engage in productive error detection and override due to reasoning biases. Reasoners often look for evidence to justify the output of the intuitive process 1 if they already believe it is correct (i.e., confirmation bias) [17]. For example, Danny justified Lisa’s comparison of the forces of friction by employing the mathematical expression for the maximum value of static friction that is not applicable in this case. If process 2 does scrutinize the provisional mental model and reasoning red flags are detected, then a new reasoning cycle begins by process 1 suggesting a new provisional mental model. The cycle repeats until a satisfactory answer is reached. In summary, to catch a mistake, a reasoner must engage in process 2, detect reasoning red flags, possess strong enough relevant knowledge to generate plausible alternatives, and assess their validity.

III. MOTIVATION AND STUDY DESIGN

The DPToR outlines cognitive mechanisms responsible for productive and unproductive reasoning pathways. In this study, we probed under what conditions students are more likely to reason productively and what factors may impact their reasoning approaches. We designed a longitudinal 8-week study in an introductory calculus-based mechanics
course. Students participated in 4 assignments spaced by ~10-14 days and administered in a web-based format outside of class. Each assignment focused on a task similar to that included in the opening paragraph. Tasks for each assignment are shown in Fig. 3. Assignment 1 included task 1 and was administered after all relevant instruction on forces and Newton’s laws. Assignments 2-4 included tasks 2-4, respectively, followed by scaffolded intervention.

The task from the opening paragraph is task 1. In task 2 a magnet weighing 10N is at rest on a refrigerator door while a hand supports the magnet from below with a 6N force; students determine the force of friction between the magnet and the door. In task 3, two pancake-like objects of different surface areas but the same mass fall to the ground after reaching terminal speeds; students compare the forces of air resistance on each object. In task 4, two identical blocks are at rest on different springs; students compare the forces on each block by a spring [18]. All tasks require the application of Newton’s 2nd law to recognize that \( \mathbf{F}_{\text{net}}=0 \) on each object, and therefore: in task 2, the force of friction between the magnet and the door points in the direction of the force by the hand; in task 3, the force of air resistance (\( F_{\text{air}} \)) is equal to the weight and therefore \( F_{\text{air}}=F_{\text{weight}} \); and in task 4, the force by the spring (\( F_{\text{spr}} \)) is equal to the weight so that \( F_{\text{spr}}=F_{\text{weight}} \).

Many students gave incorrect answers, just like Lisa and Danny. On task 2, many reasoned that the force of friction “opposes the force by the hand” and must point downward. These students typically do not recognize the need to draw a free-body diagram and include the weight of the magnet or use Newton’s 2nd law. When pointed to such an omission, many argue that they “forgot about the gravity.” On task 3, many stated that the object with a larger surface area experiences a greater force of air resistance. Such responses often contain the mathematical expression \( F_{\text{air}}=\rho\Delta v^2/2 \) to justify the dependence of \( F_{\text{air}} \) on the surface area \( A \). Students engaged in this line of reasoning neglect to recognize that this approach includes an inappropriate assumption that the objects move with the same terminal speed \( v \). Similarly, many responses to task 4 are based on the salience of the different heights of the blocks on the springs. Students think that the different spring compressions, \( \Delta x \), signify different \( F_{\text{spr}} \) and often justify this thinking with Hook’s law while inappropriately assuming identical springs (\( k_1=k_2 \)).

Through the lens of DPToR, it could be argued that student responses of this nature stem from incorrect provisional mental models cued by salient features of the tasks (e.g., different surfaces). Researchers argue that since the output of process 1 is subconscious and automatic, two approaches may be employed to improve performance. First, develop instruction focusing on a more accurate output of process 1 [19]. If the relevant knowledge is strengthened to the level of automaticity (as is often the case for physics instructors), then reasoners are more likely to immediately and subconsciously recognize its applicability correctly. Second, focus instruction on the more productive engagement of process 2 in error detection and override. Students should be able to recognize reasoning red flags and examine the validity of their provisional mental models by checking for consistency with more fundamental knowledge (e.g., Newton’s laws). In our study, we create DPToR-informed learning opportunities by 1) providing systematic spaced practices and 2) implementing scaffolded interventions for more productive process 2 engagement.

Scaffolded interventions were included in assignments 2-4 and followed each task as shown in Fig. 3. The interventions prompted the students to 1) consider alternative reasoning approaches, 2) apply relevant formal knowledge to help choose between alternative solutions, and 3) reconsider the initial response, if necessary. Below we focus on an intervention for task 4. The interventions for tasks 2 and 3 are based on similar principles.

First, students were asked to consider two expressions \( (F_{\text{spr}}=k_1\Delta x, F_{\text{spr}}=mg) \) and determine which expression(s) must be used to compare the magnitudes of the forces on each block by the spring. This question was designed to nudge students to consider alternative reasoning approaches and, if appropriate, reject the reasoning based on the assumption that \( k_1=k_2 \). To further facilitate the error detection and override, students were asked to determine which one of the choices in Fig. 4 represents the correct free-body diagrams for the two blocks. This question was designed to make the information about the blocks being identical \( (i.e., \text{equal } W) \) more salient, thus prompting students to balance \( W \) and \( F_{\text{spr}} \) according to Newton’s 2nd law and arrive at a correct answer. Finally, the students were asked whether they still agreed with their initial responses to task 4 and to elaborate on any changes in their reasoning.

A classroom discussion led by an instructor followed each web-based assignment, as shown in Fig. 3. Students considered a task from the assignment again (3rd attempt),...
discussed their reasoning with peers, and submitted their individual answers via a classroom personal response system. The instructor then facilitated a discussion converging on a normative response. An assessment task shown in Fig. 3 was included on the course exam. Students considered two identical blocks hanging from springs and compared the forces by the springs on each block.

IV. PRELIMINARY RESULTS AND MOTIVATION FOR FURTHER INVESTIGATION

One common limitation of longitudinal studies is a reduced student response rate. In our case, 39 out of 60 students enrolled in the course completed all the assignments, reducing the sample size to ~2/3. Nevertheless, this potential selection bias provides an upper bound on our results since it could be argued that those students who completed all the assignments may be more motivated to receive a higher grade in the course.

As stated above, we designed this study to probe to what extent systematic spaced practices improve recognition of the applicability of Newton’s 2nd law to novel situations and improve recognition of reasoning red flags that may lead to error detection and override. The expected desirable outcome was a higher success rate on each consecutive task.

The results in Table I show no clear improvement trajectory on the four tasks. The maximum success rate barely exceeds 50%. The scaffolded interventions do not appear to engage students in error detection and override successfully since only a few students improved on the 2nd attempt. The largest (but modest) improvement was observed between task 4 and the course exam assessment.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Correct (1st attempt)</th>
<th>Correct (2nd attempt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>41%</td>
<td>-</td>
</tr>
<tr>
<td>Task 2</td>
<td>20%</td>
<td>33%</td>
</tr>
<tr>
<td>Task 3</td>
<td>54%</td>
<td>56%</td>
</tr>
<tr>
<td>Task 4</td>
<td>46%</td>
<td>49%</td>
</tr>
<tr>
<td>Assessment for task 4</td>
<td>67%</td>
<td>-</td>
</tr>
</tbody>
</table>

We examined performances on task 4 and the exam to gain further insights into student reasoning patterns. On tasks 1-3, ~72% of students answered at least one of the tasks correctly, demonstrating their abilities to recognize the applicability of Newton’s 2nd law to a novel situation that presents reasoning challenges. This provides some evidence that these students possess relevant knowledge and skills to solve task 4 correctly as well. Nevertheless, only ~60% of these students responded correctly to task 4 after the intervention. This leads to two hypotheses. First, students who possess relevant knowledge, but answer task 4 incorrectly, may have a higher tendency to jump to conclusions without engaging in process 2 (i.e., have a lower tendency toward cognitive reflection). Second, the relevant knowledge is not simply present or absent. Instead, to reason productively, it must be instantiated to a greater depth, which may facilitate automatic recognition of its applicability (productive output of process 1) and/or increased confidence during error detection and override (productive engagement of process 2).

To test hypothesis 1, we used the cognitive reflection test developed and widely used in cognitive psychology and beyond [13–16,20]. The test consists of 3 items that cue intuitively appealing but incorrect responses that could be easily confirmed (or rejected) upon only brief reflection. For example, the first CRT item poses the question: “A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?” A solution based on basic arithmetic yields 5¢. Many, however, give a quick response of 10¢ without checking for its validity. A correct answer to each CRT item is assigned 1 point. Scores 2 or 3 indicate a stronger tendency to mediate intuition with analytical thinking.

To test hypothesis 2, we created a variable that indicates how many tasks 1-3 a student answered correctly after an intervention (upon 2nd attempt). In the following discussion, this variable, called Strength of Knowledge, provides a rough estimate of the level of knowledge instantiation. For example, a score of 1 indicates that a student not only possesses relevant knowledge but also was able to recognize its applicability to a situation eliciting intuitively appealing responses at least once. The higher the score, the deeper the knowledge is instantiated.

Since the most significant improvement in student performance occurred on the exam after an instructor-led classroom discussion, we explored how the shifts in performance between task 4 and the exam are linked to the Strength of Knowledge and a CRT score.

V. RESULTS AND DISCUSSION

The histogram in Fig 5a suggests no significant relationship between the performance on task 4 (labels C and I indicate correct and incorrect responses, respectively) and a CRT score. While more students with CRT=3 answered correctly and all students with CRT=0 answered incorrectly, the distributions of correct and incorrect responses for CRT=1 and CRT=2 are roughly the same. We used logistic regression analysis to formally verify this claim. Logistic regression is robust for a sample with more than 10 events (i.e., number of correct responses) per predicting variable (i.e., CRT score) [21]. The model for the probability of a
correct response on task 4 as a function of a CRT score suggests that the CRT score is not a statistically significant predictor of success on task 4 ($\beta=0.6, p=0.07$) [15].

Even though our data do not support hypothesis 1, it does not mean cognitive reflection skills are irrelevant to productive reasoning. A replication study with a larger sample size is needed to verify the result. It is also possible that the strength of knowledge is a more powerful predictor in cases of systematic spaced practices. Indeed, a histogram in Fig. 5b does suggest a link between student performance and the strength of their knowledge. Students who consistently answered tasks 1-3 correctly had a 100% chance of correct response on task 4. Students with a knowledge score of 2 (or 1) were slightly more (or less) likely to answer task 4 correctly, and the students with a knowledge score of 0 were very unlikely to do so. The logistic regression model for the probability of answering task 4 correctly as a function of the Strength of Knowledge variable suggests a strong statistically significant relationship between the two variables ($\beta=1.4, p=0.002$), thus supporting hypothesis 2.

Finally, we examined the relationship between student performance on the exam, task 4, and a CRT score. The results suggest that nearly all students who answered task 4 correctly (bottom row in Fig. 6) also arrived at a correct answer on the exam. About half of the students who answered task 4 incorrectly (top row in Fig. 6) recovered on the exam and gave a correct response. There does not appear to be a dependence of the shifts in student performance between task 4 and the exam on the student CRT scores. As evident from the top row in Fig. 6, students who improved their reasoning on the exam were equally likely to do so regardless of their CRT score (except those with CRT=0 who consistently underperformed). The logistic regression analysis supports the conclusion that the performance on task 4 is a strong predictor of success on the exam ($\beta=2.3, p=0.007$) while a CRT score is not statistically significant.

We argue that the students who possess relevant knowledge do not always apply it successfully because intuitively appealing responses often overshadow its applicability. If students feel confident in their provisional mental models, they do not tend to recognize the need to apply formal knowledge to scrutinize intuition-based responses. This is consistent with prior findings that novices tend to compartmentalize their knowledge instead of reasoning from fundamental principles [22,23]. Many students learn how to apply Newton’s 2nd law to solve a variety of more computationally demanding problems but struggle to recognize how to use the same knowledge as a criterion that needs to be satisfied when validating or rejecting a response.

A classroom discussion incorporating peer-peer and instructor-student interactions helped some students to transfer correct reasoning to the situation presented on the exam. However, it is still an open question whether students will be more successful in applying this knowledge to similar tasks in different contexts (e.g., comparing buoyant forces on two identical blocks floating on surfaces of different liquids at different levels of depth).

VI. CONCLUSION

We employed DPToR as a guide for developing instructional interventions to improve student reasoning on tasks that elicit intuitive incorrect responses. Analysis revealed that intuitive thinking has a strong hold on student reasoning even after systematic spaced practices. We examined factors that may impact student performance. In prior studies, cognitive reflection skills have been linked to productive reasoning on similar tasks. In this study involving systematic spaced practices, however, CRT scores do not appear to be a predictor of success. The strength of knowledge, measured by the success rate on similar tasks in different contexts, improves knowledge transfer to a different context. A classroom discussion also appears to facilitate a more productive application of relevant knowledge to a similar context. The improved performance may be attributed to two mechanisms: a strengthened recognition of the applicability of relevant knowledge and an improved recognition of reasoning red flags cued by a familiar context. However, more targeted instruction is needed to help students recognize how to apply fundamental knowledge as a criterion for validity checking. A replication study with a larger sample size is necessary to examine the validity of our findings and expand to different contexts.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under the grants Nos. DUE-1821390, 1821123, 1821400, 1821511, 1821561.
References:


Students Stumble With Inconclusive Results: An Exploratory Analysis on How Students Interpret Unexpected Results

Joss Ives, Aaron M. Kraft, James Day, and D. A. Bonn
Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC, V6T 1Z1

Success in inquiry labs often requires students to grapple with results that contradict their expectations. Previous work has shown that students who see the goal of the lab in terms of model confirmation rather than model testing struggle to engage in inquiry. Our study set out to extend this previous work by looking at the impact of asking students to hypothesize in a first year lab activity designed to produce unexpected results. Our exploratory analysis shows that hypothesizing does not play a major role in students interpreting their results correctly; instead, inconclusive results are the most significant factor in explaining correct student interpretations. We will show that these misinterpretations are not the result of model confirmation bias, but rather, misunderstanding the statistical nature of the results.
I. INTRODUCTION

The first year lab sequence at our institution teaches students experimental skills through a scaffolded scientific inquiry process [1]. By iteratively making measurements, quantitatively comparing these measurements and then reflecting on this comparison, these labs have been shown to teach critical thinking skills [2]. In the second week of this ten week lab sequence, students are introduced to the t-score, a continuous scale allowing for quantitative comparisons between two measured values [3]. The t-score is framed as a metric quantifying how distinguishable two measured values with uncertainties are. Students are given instructions shown in Fig. 1 for information on how to interpret the t-score which is formally defined as:

\[
t' = \frac{|A - B|}{\sqrt{(\delta A)^2 + (\delta B)^2}}.
\]

Reflecting on this t-score is key to the iterative inquiry in these labs.

In the third lab of this sequence, the students use this t-score to compare the period of a pendulum at 10° and 20°. Many students come into this lab unaware of the small angle approximation used to derive \(T = 2\pi\sqrt{L/g}\). The results of this lab often surprise students because a high precision measurement will uncover a breakdown in this model. These labs with unexpected results are thought to be useful for developing inquiry and modelling skills, because they force students to grapple with outcomes that contradict their own internal models [4, 5]. Forming scientific questions in response to these confusing results is a key aspect to scientific inquiry known as problematizing [6]. Therefore providing students with the opportunity to grapple with these results and supporting their success in this process is vital for the development of their scientific inquiry skills.

Previous work has shown that students struggle to problematize in these labs because they see the goal of the lab in terms of verifying rather than testing models [7–11]. This framework leads students to think that there is a specific outcome of the lab that they must achieve, leading them to ignore results that contradict that outcome. But in order to effectively grapple with contradiction, students must first interpret comparisons. In this example, interpreting their t-score correctly is necessary for problematizing. We set out to investigate how these model confirmation frameworks affect students’ t-score interpretations, focusing on the role of hypothesizing. Our main research questions were:

1. How does asking a student to hypothesize impact their ability to interpret their t-score correctly?
2. How does a student’s self-reported surprise at the result impact their ability to interpret their t-score correctly?
3. What other factors impact their ability to interpret their t-score correctly?

While our research questions and motivations were focused initially on model confirmation bias, we found that this did not play a major role in explaining correct t-score interpretations. Instead, we found that inconclusive results lead students to misinterpret their results because of an underlying misunderstanding of the statistical nature of the t-score.

II. EXPERIMENTAL CONTEXT AND METHODS

This study is an exploratory analysis of results from a first year physics standalone lab course at a large R1 institution in Canada. Our participants were students enrolled in the fall semester of this course. These students fall primarily into two groups, those taking enriched first year physics, and those enrolled in a first year science cohort program. Data were collected from the third lab in the sequence, the pendulum lab. In order to test the effects of asking students about their experimental expectation, the six sections of the course were split into three “Hypothesizing” sections and three control sections. At the beginning of the lab, each section was asked to complete a short “Start of Lab Survey” for completion credit. The control sections were asked conceptual questions on the lab while the experimental sections where asked to record their expectation. At the end of the lab all sections were asked to complete two reflection questions:

1. Based on your results in this lab, by how much did your periods at 10 degrees and 20 degrees disagree with each other? (Strong disagreement, slight disagreement, values are in tension, we can’t see much difference, or values are identical)
2. Is this the result you expected or did it surprise you?

In order to answer our three research questions, we coded the reflection questions for students’ experimental conclusion (Agreement, Disagreement, Tension), whether they used their t-score to justify this conclusion, whether they reported being surprised by the result of the experiment, and their hypothesis. The hypothesis was coded because many students provided it when discussing why they were or were not surprised, but it was not explicitly asked for in the control group. Because the information was not volunteered by all students, it was coded with three categories, Agreement, Disagreement and Unsure. Their t-scores were also collected from their lab notebooks. Their stated experimental conclusion (Agreement, Disagreement, Tension) was compared to their final t-score to come up with a binary correct interpretation variable.
that became the focus of the analysis. Table I shows the frequency at which students draw different specific experimental conclusions based on their final t-scores.

In addition to data specific to the pendulum lab, we collected responses to the Physics Lab Inventory of Critical Thinking (PLIC) [12] at the beginning and end of the semester. Incentivized with 5% extra credit each and with response rates of approximately 70%, these pre- and post-surveys assessed how well students used data and evidence to make decisions in an experimental physics context. The PLIC post semester survey also included demographic questions. For our analysis, demographic variables class standing and gender were transformed from categorical to binary variables: first year vs non-first year and under-represented vs over-represented, respectively. These two variables, as well as the PLIC pre- and post- scores, were multiply imputed using default predictive means matching in the MICE library in R [13].

The binary outcome variable for this analysis was drawn from Table I and described if their experimental conclusion matched the correct conclusion that should be drawn from their t-score. In that table, entries on the diagonal are considered correct because the stated experimental outcome matches the t-score, entries on the off diagonals are incorrect.

The students in our lab worked in groups of two or three so their t-scores and reflection question responses are correlated. Typically such correlations are accounted for using a mixed-effect logistic regression [14], however, small group sizes can lead to issues with bias and separation in these models [15–17]. Therefore, we chose to use Generalized Estimating Equations (GEE) with a logistic link function. While GEE provides similar information to the more standard logistic regressions, the coefficients are not exactly the same as they are averaged over the population. GEE also allows you to specify a correlation structure which correlates coefficients for the students within each group.

GEE is a population average modelling technique that accounts for correlated responses like those with group clustering or longitudinal studies [17–19]. It works by specifying an estimating equation for the coefficients associated with each predictor in our model. As with a logistic regression, our model is given by a logistic function where the probability of a correct/incorrect (0 or 1) interpretation is a function of the predictors \((x_1 \cdots x_m)\) and their associated coefficients \((\beta_1 \cdots \beta_m)\)

\[
P(y = 1 \text{ or } y = 0) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m)}}.
\]

Unlike logistic regression, which uses a likelihood minimizing technique to estimate the coefficients, GEE uses an algorithm known as a sandwich estimator to determine the coefficients [17, 18]. It does this in a way that accounts for correlations within groups in the data.

Ten variables were used as potential covariates for the modelling. Model selection was performed using the Quasi Information Criteria (QICu) for GEE. These metrics calculate the quality of fit and add a penalty for the number of covariates included. Covariates are added to a base model in a stepwise manner and only included in the final model if they lower the metric by 2.0. [18]. Our ten covariates were: final t-score*, categorical; hypothesizing*, binary; surprise*, binary; experimental expectation, categorical; use of t-score in reflection question answer, binary; gender, binary; class standing, binary; PLIC pre-score*, continuous; PLIC post-score, continuous; first round t-score; categorical.

Covariates indicated with (*) were used as the base model. Final t-score, Hypothesizing and Surprise were chosen for the base model because of their relationship to the three research questions with the inclusion of Hypothesizing answering question 1 and the inclusion of surprise answering question 2. The other covariates were included to answer our third research question by looking at what other factors impact students. PLIC Pre-score was chosen because it helps control for students ability to make decisions based on data in a physics context. Dummy coding was used for the final t-score and first round t-score with \(t > 3\) used as the baseline, weighted effect coding was used for the experimental expectation [20]. Because these labs stress the iterative process of experimental physics, all students are required to complete at least two rounds of experimentation.

### III. RESULTS

Our Results, shown in Table I show that nearly 80% of students did interpret their t-score correctly, but other trends are visible. The majority of the results showed that the periods of the pendulums are distinguishable, as expected, however, 10% of those students concluded that the periods agreed with each other. More striking, however, is that 45% of students who should have reported inconclusive results \((1 < t' < 3)\) instead characterized their results either as agreement or disagreement. Student reflection questions and histograms of the t-scores provided no evidence that students with t-scores close to the boundary took this into account in their analysis. A McNemar’s Chi Squared test was performed to investigate the asymmetry in this table. The test shows that the asymmetry is significant, \(\chi^2(3, N = 215) = 25.4, p = 1.2 \times 10^{-5}\). This indicates that the 20% of students who misinterpreted their t-scores are not randomly distributed and there may be difference in student interpretation based on t-score. To investigate this further, we used GEE as described above.

| Table I. Table showing the stated experimental conclusion vs the t-score from the experiment. |
|---------------------------------|-----------------|-----------------|-----------------|
| Agreement (N=54)               | \(t' < 1\) (N=38) | \(1 < t' < 3\) (N=56) | \(t' > 3\) (N=121) |
| Tension (N=36)                 | 32              | 9               | 13              |
| Disagreement (N=125)           | 4               | 31              | 1               |

}\[135\]
TABLE II. Table showing the results from Generalized Estimating Equations with effect sizes. Effect size was calculated by converting the log odds ratio to a Cohen’s d equivalent. ** indicates $p < .05$.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>log(Odds Ratio)</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final $t'$ : Tension</td>
<td>-2.20 ± .45**</td>
<td>-1.21 ± .25</td>
</tr>
<tr>
<td>Final $t'$ : Agreement</td>
<td>.06 ± .64</td>
<td>.06 ± .06</td>
</tr>
<tr>
<td>Hypothesizing</td>
<td>.75 ± .49</td>
<td>.41 ± .27</td>
</tr>
<tr>
<td>Surprise</td>
<td>.60 ± .36</td>
<td>.33 ± .20</td>
</tr>
<tr>
<td>Use of $t'$ in reflection question</td>
<td>.270 ± .58**</td>
<td>.149 ± .32</td>
</tr>
<tr>
<td>PLIC Pre-score</td>
<td>.33 ± .23</td>
<td>.18 ± .13</td>
</tr>
</tbody>
</table>

Results from GEE are shown in Table II where an exchangeable correlation structure was used. The exchangeable correlation structure was chosen based on best practices for clustered data as well as QIC comparisons between independent and unstructured correlation structures [18]. Cohen’s d is also shown as a familiar representation of effect size. Of the ten potential covariates, only one was included in the model selection process using the QICu. That only one of the six potential covariates were added in the model selection process is, in itself, interesting. Gender and Class Standing, the two demographic variables, were not included in the model. Previous work [21], suggests that gender can play a major role in how students experience first year labs where they have to work in groups, but that is not the case with the cohort in this lab. The subjects are mostly first year students, so the class standing variable may not provide much meaningful information. In future analysis on the next semester’s data, this may change because the course has a higher proportion of upper year students. The QICu similarly did not select the experimental expectation. This variable is not taken directly from the students’ hypothesis so it is not directly related to a research question. The design of our study means that only half of the students have an actual hypothesis or experimental conclusion recorded, however, we were still able to determine some of this information from the reflection questions. When asked whether their results surprised them, many students volunteered a hypothesis that was then used to code the experimental expectation variable. We believe this variable serves as a reasonable proxy for an actual hypothesis. When looking at students in the experimental hypothesizing group, 20 of 24 students who were coded with “Agreement” for their expectation recorded the same hypothesis and 23 of 27 who were coded with “Disagreement” recorded the same hypothesis. Though a proxy for the hypothesis, students’ omission of this information from the reflection question explains why it was not selected in the QICu process.

Table II shows the log odds ratios (coefficients from GEE) for the covariates included in the modelling. The log odds ratios allow us to answer our three research questions. The large coefficients for final t-score in Tension and use of t-score in reflection questions shows that these are the most significant factors explaining why students misinterpret their t-scores, answering research question three. Cohen’s d shows that both of these factors have an effect size that would be considered very large. While hypothesizing does have a positive effect, Cohen’s d shows that it’s close to a medium effect but still on the small side and less important than the t-score, answering our first research question. Our second research question is answered by looking at the coefficient for Surprise, which is even smaller than that of Hypothesizing, indicating a small effect. The most significant pedagogical factor is instead the impact of inconclusive results. Including the code that indicates whether students used their t-score to justify their reflection question allows us to distinguish between students who are misinterpreting their t-score and students who don’t even know to use a t-score. With the inclusion of this variable we can interpret the final t-score variable knowing that it’s providing information for students who actually misinterpreted rather than ignored their t-score.

Our results clearly show that inconclusive data provide a serious stumbling block for students. The odds ratio from our GEE analysis indicates that for an average student who obtained a t-score in tension ($1 < t' < 3$), the odds of interpreting that t-score correctly are 9 times lower than an average student who obtained a t-score showing distinguishability. This is a pedagogically significant difference as confirmed by the very large Cohen’s d of 1.21 ± .25. The coding for this categorical variable directly compares students who obtained a t-score in tension to those who obtained $t' > 3$, this was done because having a baseline with the largest population is a good practice and obtaining $t' > 3$ is the outcome we hope our students achieve. When changing the baseline to $t' < 1$ the log odds ratio for tension becomes $-2.26$ so the conclusions of this analysis remain the same.

We also compared GEE with an independent correlation structure to the logistic regression and confirmed that they gave the same results as expected. The independent correlation GEE model and the logistic regression model had the same covariates and same coefficients after running the model selection process.

A. Discussion

To better understand why these inconclusive results present such a challenge to students, we ran a focus group asking questions related to how they think about their t-scores. This focus group was run in Spring of 2023 after the 4th lab of the semester. There were three participants in the focus group, two of whom intended on majoring in physics. When asked how they felt about getting a t-score in tension, or any inconclusive result, a student responded: “that actually makes us uncomfortable because I may be going on a stretch, but the way we design physics experiments is to answer the question in terms of either, yes, this happens or no, it doesn’t happen. We don’t have a possibility for ‘I don’t know’, or ‘maybe’ due to these questions... Yes or no is the result. It’s either yes or no. Like those are results. Doesn’t matter if it’s yes or no,
but if it doesn’t come to that point, if it’s inconclusive, then that’s simply not a result of it.”

This thinking represents a misunderstanding of the statistical nature of the t-score and is instead indicative of the point paradigm approach to measurement and uncertainty [22]. The student believes that there is a true result of the lab and that result is either a binary yes or no. The values agree or disagree, an effect exists or does not exist. In this framework, a t-score in between one and three is not a result that can be reported. This may explain why students who end the lab with inconclusive results are much more likely to misinterpret their results. It’s important to note that not everyone in the focus group shared this understanding as another student suggested that they thought inconclusive results were still valid as they were just on the spectrum between being confident in a result and not confident in a result. While the expert sees the t-score as a continuous spectrum with a grey area in the tension region, a novice may see it as three distinct categories. Agreement and Disagreement exist on the opposite ends of the same spectrum, but tension exists somewhere else entirely as it’s not a viable result of the lab.

This thinking is further explained by another student in the group who, when asked if they would interpret a t-score of 2.9 differently from a t-score of 3.1, said that those results were completely different because “The way we’ve been instructed is just to think of it that way, like it’s bigger than three than... we’re confident that it’s different. That is smaller than three, regardless of how smaller, it’s uncertain.” This student is suggesting that they were told to follow the cut-offs in t-score as hard and fast rules and not think of them as a spectrum. While this instruction is inconsistent with how t-score is presented in our labs, this student felt that the instruction suggested distinct cut-offs because of the accidental bias in the slide presenting the t-score shown in Fig. 1. This kind of thinking may result in the misinterpretations we have observed.

These results provide a clear path forward to better support students nuanced approach to statistical thinking. We can draw two main lessons from this study, first, students need to end our labs with conclusive results, and second, we need to adjust our instructions for the t-score to support expert-like statistical thinking. The first point is relatively easy to address. In the case of our pendulum labs, students are given the tools to perform high quality conclusive measurements, but they sometimes lack enough time. This study shows that these time crunches are not an insignificant factor in student success. With this knowledge in mind, we are now piloting a two week pendulum lab to ensure that students have enough time to complete a high quality conclusive experiment. Broadly speaking, this also points to the importance of providing students with adequate tools in inquiry labs. We must be mindful of the measurement quality needed to achieve learning goals in these labs and ensure the tools provided can achieve that quality. For example, students need to measure the period in our lab with a relative uncertainty of 0.1% to observe the small angle approximation. We are careful to design the experiment such that they can achieve this precision with just a stopwatch and repeated measurements.

The t-score was initially introduced to support expert-like statistical thinking and overcome the point paradigm [3]. And while it has been extremely successful in building scientific critical thinking skills, more work is needed. The third student in the focus group shows us exactly the challenge we face in our current t-score framing when they say that they have “just got to trust the statistics and the studies of other people that this [the categorical t-score framing] is like good absolute values that you have to follow.” In other words, our current instruction is inadvertently biasing students towards this framework and does not support students when their results are inconclusive. To overcome this, we are reworking the t-score instructions to emphasize the continuous nature of the scale and the validity of inconclusive results. The next step of this study will present the new instructions and assess their impact.

IV. CONCLUSION

We investigated what factors contribute to students misinterpreting their results in a first year lab with unexpected outcomes at a large R1 institution in Canada. Motivated by previous work on the role of model confirmation frames in these types of labs, we designed a study to look at the impact of asking students to hypothesize. Our exploratory analysis shows that hypothesizing had a very small positive impact but was likely not pedagogically significant. Instead, our results clearly show students struggle the most when obtaining inconclusive results. Our GEE analysis shows that for students who obtain an inconclusive result, the odds of characterizing their t-score correctly are 9.06 times lower than those who obtain a conclusive result, with a Cohen’s d of -1.21 indicating a large effect. Focus group interviews show that these students may struggle because they do not understand that an inconclusive result is still a result. These focus groups also uncovered further issues with how students interpret their t-scores. Finally, we’ve proposed changes to how t-scores are introduced in our lab and will investigate this intervention in a follow up study.

ACKNOWLEDGMENTS

A.M.K. designed the study, collected the data, analysed the data, and wrote the manuscript with supervision from J.D. J.I. and D.A.B.


Research on a faculty support program for working with learning assistants

Adrian Juanson and Gina M. Quan
Department of Physics and Astronomy, San Jose State University, San Jose, CA, 95192-0106

Jennifer S. Avena
Department of Biological Sciences, San Jose State University, San Jose, CA 95192

Alexandra Gendreau Chakarov
Department of Computer Science, San Jose State University, San Jose, CA 95192

Learning Assistant Reflection, Guidance, and Exploration (LA-RGE) is a new program at San Jose State University designed to provide support for university faculty working with learning assistants (LAs). Faculty in LA-RGE attend facilitated bi-weekly meetings throughout the semester which discuss various topics about pedagogy, LAs, and equity. In this presentation, we will discuss a research study on what kind of support is being provided by this program and how that can potentially lead to better partnerships between faculty and LAs. During one-on-one interviews, we asked faculty to discuss their personal experiences over the time they have worked with LAs, and if LA-RGE meetings have had any effect on their perspective and relationship with their LA. We then categorized the different effects and support that faculty were receiving through LA-RGE. This work will be important toward understanding how support for faculty can lead to a more productive partnership between faculty and LAs.
I. INTRODUCTION

The learning assistant (LA) model is a well-established model for creating a better learning environment for students. Learning assistants are undergraduate students who help university instructors facilitate small group discussion within the classroom, typically in interactive and reformed classrooms [1]. By keeping students engaged, promoting productive group work, and leading students into an understanding of the content, LAs help students experience more learning gains through these transformed classes [2]. LAs also bring a unique perspective into the classroom, one that serves as a bridge between that of a student and an instructor [3]. As undergraduate students, LAs may not have experience in being an educator nor formal training in education. To help LAs become effective facilitators, LAs engage in pedagogy training concurrent with their experience in the classroom. While pedagogy trainings vary institution to institution, they typically discuss questioning strategies, metacognitive techniques, and classroom equity [4, 5].

In addition to pedagogy support for LAs, we argue that support for instructors is necessary to fully incorporate their LA into their own course. Because of this, the learning assistant program at San Jose State University has developed the Learning Assistant-Reflect, Guide and Explore program (LA-RGE). LA-RGE acts as a parallel to the LA pedagogy course at San Jose State University, where instructors can learn and share new ideas for how to better incorporate LAs within their own classroom, as well as develop strategies for partnering with LAs. In this study, we will explore the different benefits that faculty receive from LA-RGE. We present an emergent framework that describes these different benefits. This work will help us gain a better understanding of ways to improve support for faculty working with LAs.

II. CONTEXT

San Jose State University is located in San Jose, California. The school is a Minority Serving Institution (MSI) where roughly 30% of students are first-generation college students. The Learning Assistant Program was formed in 2016, with the purpose of supporting active learning in transformed classes in the College of Science. While the program was initially located solely in the College of Science, it is now jointly supported by the College and Peer Connections, with additional support from the CSU CREATE Program. Peer Connections provides a number of peer-educator services (e.g., mentorship, course-embedded tutoring, workshops) and is housed within Student Affairs. First-semester LAs enroll in a 3-unit pedagogy course, which provides a supportive peer community for developing their teaching practice. All LAs are also expected to meet for a weekly planning meeting with their faculty partner who is the instructor of the discipline-specific course.

The LA program at San Jose State University has been experiencing a period of growth. In the 2022-2023 academic year, it served courses in biology, chemistry, computer science, computer engineering, and physics. In Spring 2023, the program hired 30 LAs, supporting nearly 2,000 students in 35 course sections. To support this growth, the program developed two supports for faculty working with LAs: LA-RGE (the focus of the paper) and a summer half-day retreat. The retreat was jointly run by San Jose State University and San Francisco State University in 2022 and served 21 faculty.

A. LA-RGE Background/History and Goals

While LA Programs have historically focused efforts on supporting LAs and students, relatively little is known about effective support for instructors with LA-supported courses. In order to support instructors using LAs in integrating active and inclusive teaching practices into their classrooms and in fostering partnerships with their LAs, in Spring 2022, the LA Program at San Jose State began to offer faculty the opportunity to engage in a faculty learning community termed the Learning Assistant-Reflect, Guide, Explore Program (LA-RGE). This initial iteration was co-facilitated by two program directors and the LA pedagogy course instructor. We found this facilitator composition a valuable asset to the development of LA-RGE meetings, and thus we continued this structure of co-facilitation by one experienced lead facilitator and the current LA pedagogy course instructor in this past academic year. All instructors and graduate teaching assistants who were actively using LAs or who had used LAs in previous semesters were invited to participate in LA-RGE. Each cohort of LA-RGE lasts one semester, and faculty can continuously participate in them over multiple semesters. Eight instructors and TAs chose to participate in Fall 2022, and nine chose to do so in Spring 2023. LA-RGE typically met every other week, for a total of eight meetings per semester, and instructors were compensated for their time.

For the past year, the objectives of LA-RGE were as follows: (1) Faculty will reflect on STEM education teaching practices, which include understanding how learning happens and refining skills to facilitate student discussion and reflection with LAs, (2) Faculty will receive guidance and information that will empower them to make informed decisions on how to modify their practices and integrate new approaches to their teaching practice, (3) Faculty will explore what their LAs are learning about pedagogy and hear from other instructors who have taught with LAs or who currently teach with LAs, (4) Faculty will identify ways to foster partnerships with LAs to leverage their knowledge as students in creating and supporting an active, inclusive classroom community.

In Fall 2022, discussion topics focused on fostering LA-instructor partnerships, formative assessment, and equitable practices to support students’ sense of belonging and a welcoming classroom environment. In Spring 2023, we discussed the LA model, active learning strategies, including group work and talk moves, ways to create equitable learn-
ing environments, which included discussions around implicit bias, stereotype threat, and mindset, and ways to apply these topics to LA-instructor weekly meetings. The structure of each meeting varied depending on topic but generally included a community builder (icebreaker) followed by discussion and an activity with small groups and/or the large group; these discussions focused around a pedagogical topic, which in some instances reflected the topic discussed with LAs in the pedagogy course. This past year, we also created a space using Google Slides termed the Goals Board for faculty to intentionally create and share their goals and track progress working towards these goals related to teaching, working with LAs, and LA-RGE. Each instructor had their own slide that they could build upon and revisit at each LA-RGE meeting.

III. METHODS

A. Data Collection

Since we wanted to better understand how LA-RGE was influencing instructor’s views on LAs, we invited all instructors who have participated in the LA-RGE meetings or retreat to be interviewed. We interviewed three faculty members—all three participated in LA-RGE and two of the three had also attended the retreat. During these interviews, we followed a semi-structured interview format using an adapted protocol from a previous study that examined the different feedback faculty receive from learning assistants [6]. We added several questions targeting how instructors felt about LA-RGE specifically. Topics included whether participating in LA-RGE had an overall impact on the instructor’s interactions with LAs, if there were any changes that occurred due to the meetings, a description of the work they do with LAs, how comfortable they are working with LAs, as well as how they receive and integrate LA feedback.

B. Data Analysis

Interviews were conducted through Zoom and professionally transcribed. Our analytical process began by first identifying moments where faculty described LA-RGE. We identified particularly salient moments where faculty described how LA-RGE was influencing how faculty saw LAs or their relationships with LAs. We then developed analytic memos [7] of specific benefits that were being expressed by different instructors. Writing these analytic memos helped us refine the characterizations of the different benefits that instructors were experiencing through the LA-RGE program. Initially, we noticed that faculty described pedagogical, social, and emotional benefits, which are consistent with prior research on collaborative faculty learning environments [8]. However, we also noticed that particular benefits were tied to the structure of LA-RGE and the LA program. We then developed a preliminary framework characterizing the different ways that faculty valued LA-RGE. Our framework includes emergent benefits—goals, recognition, and partnership. We now describe each of the dimensions of this preliminary framework and provide illustrative examples from our interview data.

IV. ANALYSIS

A. Goals

We defined the “goals” category as involving the creation, revision, or tracking of instructors’ pedagogical goals. Within the structure of LA-RGE, goals are a recurring component. In nearly every meeting, facilitators will invite instructors to work on their goals board, either adding a new goal in relation to the day’s topic, marking progress toward previous goals, or evaluating whether any goal(s) need to be revised in some way. When asked what a typical LA-RGE meeting looks like for them, Adam says:

Yeah. So during the very beginning of the semester we create our own individual goals for [LA-RGE]. But it’s goals and proposed actions and if things were completed, we had a different kind of structure. We created [goals in] the beginning, and then we always start by looking back at that. I think we pretty much always have started that way, and so then we can really think of what are we trying to accomplish.

As described by Adam, instructors are given the time and space to think about what individual goals they want to achieve within the semester regarding LAs. These goals are revisited throughout the semester, and being given this time at the start of the meeting helps instructors be more intentional in thinking about what they are trying to achieve. This kind of reflection happens in both individual and collaborative settings. Adam continues to describe what it looks like for other faculty members to have a discussion about them.

So we kind of get to know other people and their goals. Yeah, so that’s what [LA-RGE meetings] kind of look like. Very discussion-based, but add that we’re revising goals throughout so that it is focused on doing something with it, not just getting ideas, but taking action as well.

Adam points out how instructors are not only creating goals within LA-RGE, but also revising them throughout the semester. Instructors will also discuss these goals with others in the meeting, getting different perspectives of others and their own goals. Adam also emphasizes the revision process as helpful toward “doing something with it,” not just getting ideas, suggesting that he finds the goals structure to be particularly helpful toward enacting those goals.

B. Social

Another aspect of LA-RGE that came up in multiple interviews, is the social benefit one gets from participating in
it. We define this category as the interactions among participants that led to a sense of community, belonging, or collective engagement. Because LA-RGE meetings primarily involve small-group and whole-group discussions, instructors get the opportunity to talk to their peers about various pedagogical topics. When asked what the most useful thing about LA-RGE was, Adam says:

I think the most useful thing is being able to just talk about these ideas with other instructors. There aren’t that many opportunities to really discuss different topics related to pedagogy and teaching in general, but having this common thing that we’re all thinking about how to use our LAs to make better, more inclusive, more interactive classes ... But we have very different structures. Some people are in larger lecture type classes, and so we’re able to do this group problem solving as a community. And so I think that builds our own sense of belonging, in that case, to this learning community.

To Adam, the most important aspect of LA-RGE is the opportunity that it presents to talk with other instructors. He points out that instructors do not have a lot of opportunities to discuss different topics in pedagogy and teaching. LA-RGE provides that opportunity and space for instructors to have those conversations, and even apply a more focused lens since the main focus of discussion is about how to better incorporate LAs into their class. We characterize this benefit as “social” because Adam characterizes the feel of their community as a “learning community” with a “sense of belonging.” In a different vein, during Cynthia’s interview she characterized the social experience of LA-RGE as being relaxing, conversational, and appreciative. When asked about what emotions she feels during the meeting she says:

Relaxation. Well, I think that there are a lot of different environments that are very high stress. The classroom is ... I’ve got to always be on my game and on my toes. In LA-RGE I feel like I can just sit back a little bit and have a conversation and be supported and appreciated for the work I’m putting in.

Cynthia describes the LA-RGE environment as a space separate from the high stress environment that the classroom can be. During these conversations with other faculty members she feels supported and appreciated through the different interactions with faculty. While Cynthia does not explicitly describe the LA-RGE community within this quote, we categorized it as social because she links the conversations she has with others with her sense of being supported and appreciated. Fully incorporating an LA into a course can be difficult, but having support from other faculty can be helpful when working through that goal.

C. Pedagogical

The “pedagogical” category refers to the sharing of knowledge, strategies, or ideas pertaining to classroom teaching. Each LA-RGE meetings’ objectives will usually focus on one or more pedagogical topics (e.g., talk moves, mindset, group work, microaggressions/microaffirmations). These topics often mirror what LAs are learning in their pedagogy class, and thinking about how instructors can also use those strategies. Topics brought forward by the facilitators are not the only mechanism for instructors’ pedagogical learning; because of the discussion-based nature of the meetings, other participants will often share their own experiences and strategies that could inspire ideas in another instructor. In an interview question asking “how would you help a new faculty member decide whether or not to participate in LA-RGE,” Barbara says:

I would just say, ‘Oh my gosh, I can’t describe how valuable [LA-RGE] is because you get ideas that would have never even occurred to you that...’ And you might think, ‘Oh, someone else has a very different context.’ [Small lab class] has a very different context than a large lecture, right? But it doesn’t matter. What happens is they share what they’re doing and you might not do exactly what they do. Instead, just the discussion like an idea sparks that’s relevant to you.

We characterize this quote as pedagogical because Barbara describes LA-RGE as a useful space where faculty members can share different practices and strategies they use in their own classrooms. Though the instructors participating in these meetings all have different classroom environments, being exposed to new ideas could be useful in inspiring how to use a strategy in the context of their own class. Continuing Adam’s excerpt from earlier, when he was asked what the most useful takeaway from LA-RGE, he stated:

[LA-RGE] gives the space to focus on these particular topics, right? Because so often instructors are focused on the content of what they’re teaching next, and maybe some of the delivery of it. But not, yeah, how to better incorporate LAs into that and how to better use my time with that.

As Adam describes, the pedagogical benefits from LA-RGE can help instructors look beyond classroom content, to consider the structure of the classroom and the integration of LAs. This benefit is supported by LA-RGE discussions which involve explicit thinking and talking about the different pedagogy surrounding LAs, as well as the implementation of these pedagogical ideas.

D. Recognition

While the previous elements of our framework have been directly tied to the goals for the LA-RGE program, one unexpected benefit was recognition—the sense of feeling seen and valued by others. During Cynthia’s interview the idea of recognition or appreciation came up, when asked what a typical LA-RGE meeting looks like for her:

[LA-RGE] gives us the opportunity to be, I think, appreciated for the work we’re doing, but also supported
in the work we’re doing. Cynthia describes how working with LAs, and especially creating a classroom structure that fully incorporates LAs, is a lot of work. Other faculty might not understand this, but instructors that are also working towards the same goal would understand the effort required to make students’ experience with LAs better. We interpret Cynthia’s quote as connecting this sense of recognition that is also tied to the pedagogical and social “support” offered by LA-RGE as well. Cynthia goes on to say:

I think that there’s a part about LA-RGE that recognizes that the things that your LAs are doing in your course are because of you. That doesn’t necessarily translate to your students, but it’s really nice to know that somebody recognizes that.

We see Cynthia’s statement that “somebody recognizes that” to refer to ways that LA-RGE facilitators and participants notice and appreciate the effort she is putting in. As she suggests, this type of recognition might not translate to her students because they don’t see the behind-the-scenes work it takes to fully incorporate an LA into a class.

E. Partnership

Our final category is “partnership” which refers to the ways in which LA-RGE supports faculty in developing productive relationships with their LAs. Another important component of partnerships is the type of feedback that is being received and given between the two individuals. LA-RGE often asks instructors to reflect on their partnerships with their own LAs, and to brainstorm how to support LAs’ implementation of different pedagogical topics. When asked if LA-RGE has influenced how they interact with LAs, Barbara says:

[Prior to LA-RGE] I’d ask [my LA] any question because they’re partners, but I didn’t understand the relationship of power and how... I guess for me, if I were in their shoes, I’d be willing to answer any question because the course is important and I understand my opinion is valuable, but I don’t think I would’ve felt the pressure that maybe some LAs might feel being asked from a professor and a person of power.

In Barbara’s quote, we see how LA-RGE helped her reflect on their relationships with LAs, in particular identifying the pressure and expectations on the LAs that can be amplified by the power differential between LAs and faculty. Barbara attributes LA-RGE to be a space where she could consider these different ideas, suggesting the meetings help instructors consider dynamics they were not considering before.

V. SUMMARY AND DISCUSSION

Because of the positive learning gains that LA programs bring, more universities are starting to implement their own version of the LA model [9]. As our own program has been growing to include a number of faculty who are new to working with LAs, we found it important to support them in an analogous way to LAs. In this paper, we focused on one of these supports—an ongoing faculty learning community called LA-RGE. Through our interviews, we were able to gather the different ways that faculty benefit from these meetings, as well as understanding how the program supported these benefits. From our analysis we developed a preliminary framework describing five categories of support: Goals, Social, Pedagogical, Recognition, and Partnership.

LA-RGE supports faculty by helping them be prepared for their future partnerships with LAs, and giving instructors the pedagogical tools to improve their understanding of how to incorporate an LA within their course. In addition, having the space to discuss these different topics within an ongoing community of faculty members helps support them through the potential challenges they might face when trying to fully incorporate an LA into their course.

We see this framework as helpful for future facilitators of LA-RGE and communities similar to LA-RGE by illustrating the value the instructors receive and what is most important to them. Gaining a better understanding of what aspects of LA-RGE leads to those benefits can support facilitators in designing for particular faculty learning outcomes. We also see this framework as potentially helpful for instructors who work with LAs, as it can help them understand the different dimensions of support they may look for in other faculty development programs they may participate in.

Finally, this work is also helpful for those doing research on LA program effectiveness. While much is known about the various benefits LAs bring to the classroom, less is known about how instructors’ implementation of LA-supported pedagogy impacts those benefits. We argue that understanding how LA programs can support faculty in more productively working with LAs can also improve the experience of LAs and students in LA-supported classrooms. To explore this in future work, more data collection is necessary to gain a more comprehensive understanding of all the benefits that come with a program like LA-RGE. As LA-RGE continues to evolve, we plan to refine the categories of this framework. Once we have a more refined, and robust framework we would like to compare it to other faculty development programs and theoretical ideas in the future. We would also like to explore how programs like LA-RGE contribute to partnerships between faculty and LAs, as well as how those partnerships impact their classrooms as a whole.

VI. ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. DUE 2234071. This work is also supported by an award from the California State University Creating Responsive, Equitable, and Active Teaching and Engagement (CREATE) Awards Program.
Rasch Analysis of the Quantum Mechanics Concept Assessment

Jesse Kruse¹ and Bethany R. Wilcox¹

¹Department of Physics, University of Colorado, 390 UCB, Boulder, CO 80309

Quantum mechanics is a subject rife with student conceptual difficulties. In order to study and devise better strategies for helping students overcome them, we need ways of assessing on a broad level how students are thinking. This is possible with the use of standardized, research-validated assessments like the Quantum Mechanics Concept Assessment (QMCA). These assessments are useful, but they lack rigorous population independence, and the question ordering cannot be rearranged without throwing into question the validity of the results. One way to overcome these two issues is to design the exam to be compatible with Rasch measurement theory which calibrates individual items and is capable of assessing item difficulty and person ability independently. In this paper, we present a Rasch analysis of the QMCA and discuss estimated item difficulties and person abilities, item and person fit to the Rasch model, and unidimensionality of the instrument. This work will lay the foundation for more robust and potentially generalizable assessments in the future.
I. INTRODUCTION & BACKGROUND

Quantum mechanics is a notoriously difficult subject to learn and understand. There has been much characterization over the past thirty years on student difficulties and misconceptions in undergraduate [1] and graduate quantum mechanics courses [2]. These include difficulties with reconciling quantum concepts and classical concepts, properties and representations of wave functions, distinguishing between three-dimensional Euclidean space and Hilbert space, measurement and expectation values, Dirac notation, and many more [1]. In addition to the complexity of quantum mechanics, instructors disagree on which topics to include in, and how to teach, the subject [3]. For example, instructors disagree on whether to present quantum mechanics in a spins-first or wavefunctions-first approach, whether to present an axiomatic or historical approach, and whether wavefunctions represent a matter wave, information wave, or something else entirely [3].

Because of the plethora of challenges students face in learning quantum mechanics and because of the lack of consensus on what and how to teach the subject, it is difficult to establish clear learning goals that are relevant across institutions. This has posed an issue for evaluating student learning with research-based assessments. If developers design a test that contains a certain number of subjects, it is possible that it won’t be applicable to classes that chose not to cover all those topics. For example, spins-first courses often have not finished solving the full Schrödinger equation for the hydrogen atom by the end of the first semester, so questions involving the hydrogen atom on an assessment won’t provide useful measures of learning for that class. Therefore, it would be advantageous if a modular assessment was developed that could accommodate the variety of instruction and learning goals inherent in undergraduate quantum physics education.

Currently, there are around ten research-validated assessments for modern physics and quantum mechanics [4–10]. These cover a variety of topics such as measurement, wave functions, Dirac notation, incompatible operators, scattering, tunneling, time dependence, and many more. However, there is still a multitude of topics that aren’t assessed by these instruments such as most topics in a second semester of quantum mechanics and some topics that may be covered in either semester like entanglement, Bell inequalities, and topics related to quantum information.

In addition, these assessments have all been validated using classical test theory. Classical test theory (CTT) consists of a few statistical measures of test scores that look at item difficulty, discrimination, reliability, and overall consistency within the test. This is the most commonly used framework for validating assessments within PER [11]; however, this approach brings with it some fundamental limitations. These include the fact that the scores you get are always dependent upon the sample used for calibration. This means the item difficulty and discrimination will be different for different samples, and the ordering of the questions can have significant effects on student performance. In addition, differences of students’ abilities in CTT don’t have a well-defined meaning whereas in Rasch measurement theory they do [12].

Rasch measurement theory (RMT) and more generally item response theory (IRT) are probabilistic models of student responses to test items that are functions of the person ability1 and item difficulty. In RMT, the base assumption is that the probability of a person answering an item correctly is dependent only on the difference between their ability and the item difficulty. This offers an advantage over CTT because the item difficulties and person abilities are computed together and share a common scale where comparison is actually meaningful [12]. The Rasch model also allows us to estimate person ability independent of which items are used and the item difficulties independent of the people used to calibrate it [13].

This paper will provide a brief overview of Rasch measurement theory, discuss the data used in analyzing the QMCA, and present our results for item difficulties, person abilities, and how well the data fit our model of measurement. We hope to add well-fitting items to a quantum physics test bank that instructors can use to design their own assessments while still generating robust comparable measures of student learning.

II. RASCH MEASUREMENT THEORY

The founder of Rasch measurement theory was a Danish mathematician named Georg Rasch. He postulated that in order for psychological measurement to truly be a measurement, the process of assessing the property under study must follow certain criteria. The numbers that we assign to the property under study must have a common scale and be comparable to one another on the interval level [12]. Rasch postulated a frame of reference which was necessary for interpreting the results for an assessment which consists of:

1. the class of items on an assessment that target the construct under study,
2. the class of persons who are relevant to be assessed
3. the conditions of the administration of the assessment.

He further postulated that if the conditions of the frame of reference are appropriate, an examinee’s performance on a given item should be dependent on two parameters: the examinee’s ability \( \theta_n \), and the item’s difficulty \( \delta_i \). These two parameters are measured on the same scale and are sometimes referred to generally as the person and item location. Rasch theorized that in order for an assessment to truly measure an underlying trait, the comparison between any two persons must be independent of the choice of items used, and the comparison between any two items must be independent of the people that interact with the items [12]. This property is known as the invariance of comparisons, and it directly leads to the probability distribution that describes students’

---

1 Note, “ability” refers to the latent trait that the statistical models quantify. Fundamentally, however, it is a measure of performance as opposed to innate ability. This term is used for consistency with the existing literature. However, this term is potentially problematic, particularly with respect to the interpretation of performance gaps between subgroups of students.
likely than not, and vice versa. The log-odds ratio which describes this behavior is given by:

\[ P(x_{ni} = 1 | \theta_n, \delta_i) = \frac{e^{\theta_n - \delta_i}}{1 + e^{\theta_n - \delta_i}}, \]

(1)

where \( x_{ni} \) is the outcome of person \( n \) answering item \( i \) correctly. In general \( \theta_n \) and \( \delta_i \) vary continuously from \( -\infty \) to \( +\infty \), but we perform a linear transformation so the mean is 0 and the standard deviation is 1. We see that when \( \theta_n \) is larger than \( \delta_i \), the probability of answering correctly is more likely than not, and vice versa when \( \theta_n < \delta_i \). In addition, the log-odds ratio which describes this behavior is given by:

\[ \ln \left( \frac{P}{1 - P} \right) = \theta_n - \delta_i. \]

(2)

When we have the same person attempting two different items with difficulties \( \delta_1 \) and \( \delta_2 \), the difference in the log-odds is simply the difference in the item difficulties \( \delta_1 - \delta_2 \). This means that the likelihood of answering one item correctly and the other incorrectly is a function only of the difference in item difficulties, not the person’s ability, thus demonstrating the invariance of comparisons [12]. Another fundamental assumption of RMT is that all of the items on an assessment are parallel, meaning they independently measure some aspect of the construct under study. The primary task of RMT is to estimate the person abilities and item difficulties, not the person’s ability, thus demonstrating the invariance of comparisons [12].

III. CONTEXT & METHODS

The Quantum Mechanics Concept Assessment (QMCA) is a conceptual assessment for evaluating student learning in upper-level, undergraduate quantum mechanics courses [4]. The QMCA consists of 38 multiple-choice items covering measurement, the time-independent Schrödinger equation, time evolution, wave functions, boundary conditions, probabilities, and expectation values. It was adapted from the free-response format Quantum Mechanics Assessment Tool (QMAT), and it has undergone many revisions and verifications of face and content validity including a CTT analysis [4], an exploratory factor analysis [14], and a modified module analysis [15]. However, there has yet to be a Rasch analysis of the QMCA or any quantum assessment for that matter.

The data used in this analysis were gathered from eight separate institutions ranging in size, research output, location, and even nationality. Seven of the institutions were located in the continental United States, but one sample was from a university in South Africa. The responses to the QMCA were gathered from the fall semester of 2018 to the spring semester of 2022, with a majority coming from the 2018-2019 academic school year. Around 30 students had incomplete attempts, so it was necessary to establish a criterion for whether or not to include them in the aggregate analysis. We decided to use a 90% completion cutoff, so if students answered at least 90% of the items, it could be considered a good-faith attempt. This corresponds to answering at least 35 of the 38 questions. Note that including them did not affect the results significantly. After combining all the data, we had a total of 403 responses, which is large enough to reliably estimate parameterizations in the Rasch model [16].

IV. RESULTS & DISCUSSION

A. CTT Analysis

We start off by looking at the CTT statistics and comparing them to published results. The definitions of the item difficulty index \( P \), item discrimination index \( D \), point biserial index \( r_{pbi} \), Kuder-Richardson 20 coefficient KR-20, and Ferguson’s delta are all discussed in [11]. Table I summarizes these results and compares them to Sadaghiani and Pollock’s study [4]. We see that the statistics calculated from our data are slightly higher above the threshold than that of Sadaghiani and Pollock’s, which may be a result of being a newer version of the QMCA.

B. Exploratory Factor Analysis

Before fitting our QMCA data to the Rasch model, we need to check whether our data are sufficiently unidimensional. The number of dimensions assessed by an instrument dictates what sort of model we should apply to it. Our initial naive hypothesis is that the QMCA assesses a single dimension of quantum mechanics proficiency at the first semester upper-level undergraduate level, but in order to confirm this we need to do an exploratory factor analysis (EFA). To do this, we used the \texttt{mirt} package in R (a multidimensional item response theory modeling package) [17] and specified the number of assumed dimensions. This gives us the factor loadings from which we interpret whether it is an adequate fit.

The factor loadings for the unidimensional EFA are in Table II. The factor loadings represent what percentage of the variance on that item can be accounted for by the assumed underlying factor. If an item has a factor loading greater than or equal to approximately 0.3 then it is said to be well described by the factor [14]. We see from the table that the items that

<table>
<thead>
<tr>
<th>CTT Stats</th>
<th>Target Values</th>
<th>Ref [4]</th>
<th>Our Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num. Stu.</td>
<td>...</td>
<td>263</td>
<td>403</td>
</tr>
<tr>
<td>Num. Items</td>
<td>...</td>
<td>31</td>
<td>38</td>
</tr>
<tr>
<td>Ave. ( P )</td>
<td>&gt; 0.3</td>
<td>0.54</td>
<td>0.55</td>
</tr>
<tr>
<td>Ave. ( D )</td>
<td>&gt; 0.3</td>
<td>0.42</td>
<td>0.45</td>
</tr>
<tr>
<td>Ave. ( r_{pbi} )</td>
<td>&gt; 0.2</td>
<td>0.35</td>
<td>0.39</td>
</tr>
<tr>
<td>KR-20</td>
<td>&gt; 0.7</td>
<td>0.76</td>
<td>0.84</td>
</tr>
<tr>
<td>Ferguson’s ( \delta )</td>
<td>&gt; 0.9</td>
<td>0.97</td>
<td>0.986</td>
</tr>
</tbody>
</table>

### Table II. Exploratory factor analysis for 1 factor. The factor loadings for each item are under F1. Factor loadings in red indicate that they are below the 0.3 threshold.

<table>
<thead>
<tr>
<th>Item</th>
<th>F1</th>
<th>Item</th>
<th>F1</th>
<th>Item</th>
<th>F1</th>
<th>Item</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.86</td>
<td>11</td>
<td>0.39</td>
<td>21</td>
<td>0.26</td>
<td>31</td>
<td>0.53</td>
</tr>
<tr>
<td>2</td>
<td>0.91</td>
<td>12</td>
<td>0.21</td>
<td>22</td>
<td>0.87</td>
<td>32</td>
<td>0.40</td>
</tr>
<tr>
<td>3</td>
<td>0.44</td>
<td>13</td>
<td>0.40</td>
<td>23</td>
<td>0.92</td>
<td>33</td>
<td>0.47</td>
</tr>
<tr>
<td>4</td>
<td>0.11</td>
<td>14</td>
<td>0.41</td>
<td>24</td>
<td>0.53</td>
<td>34</td>
<td>0.57</td>
</tr>
<tr>
<td>5</td>
<td>0.26</td>
<td>15</td>
<td>0.34</td>
<td>25</td>
<td>-0.10</td>
<td>35</td>
<td>0.49</td>
</tr>
<tr>
<td>6</td>
<td>0.72</td>
<td>16</td>
<td>0.39</td>
<td>26</td>
<td>0.17</td>
<td>36</td>
<td>0.67</td>
</tr>
<tr>
<td>7</td>
<td>0.59</td>
<td>17</td>
<td>0.44</td>
<td>27</td>
<td>0.82</td>
<td>37</td>
<td>0.15</td>
</tr>
<tr>
<td>8</td>
<td>0.65</td>
<td>18</td>
<td>0.63</td>
<td>28</td>
<td>0.57</td>
<td>38</td>
<td>0.05</td>
</tr>
<tr>
<td>9</td>
<td>0.43</td>
<td>19</td>
<td>0.66</td>
<td>29</td>
<td>0.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.43</td>
<td>20</td>
<td>0.18</td>
<td>30</td>
<td>0.49</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table III. Item fit statistics for the whole QMCA under the Rasch model. S-X2 is the signed chi-squared statistic, df is the degrees of freedom which is variable depending on the binning for S-X2, RMSEA is the root mean square error of approximation, and p is the probability of observing those data or more extreme values. The color red indicates that the RMSEA is at or above the 0.06 threshold.

<table>
<thead>
<tr>
<th>Item</th>
<th>S-X2</th>
<th>df</th>
<th>RMSEA</th>
<th>p</th>
<th>Item</th>
<th>S-X2</th>
<th>df</th>
<th>RMSEA</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51</td>
<td>22</td>
<td>0.06</td>
<td>0</td>
<td>20</td>
<td>55</td>
<td>23</td>
<td>0.06</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>22</td>
<td>0.06</td>
<td>0</td>
<td>21</td>
<td>29</td>
<td>21</td>
<td>0.03</td>
<td>0.13</td>
</tr>
<tr>
<td>3</td>
<td>42</td>
<td>22</td>
<td>0.05</td>
<td>0.01</td>
<td>22</td>
<td>44</td>
<td>22</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>34</td>
<td>22</td>
<td>0.03</td>
<td>0.05</td>
<td>23</td>
<td>80</td>
<td>22</td>
<td>0.08</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>32</td>
<td>22</td>
<td>0.03</td>
<td>0.08</td>
<td>24</td>
<td>27</td>
<td>22</td>
<td>0.02</td>
<td>0.22</td>
</tr>
<tr>
<td>6</td>
<td>26</td>
<td>22</td>
<td>0.02</td>
<td>0.25</td>
<td>25</td>
<td>118</td>
<td>22</td>
<td>0.10</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>28</td>
<td>23</td>
<td>0.02</td>
<td>0.20</td>
<td>26</td>
<td>54</td>
<td>23</td>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td>8</td>
<td>24</td>
<td>18</td>
<td>0.03</td>
<td>0.15</td>
<td>27</td>
<td>48</td>
<td>22</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>23</td>
<td>0.00</td>
<td>0.74</td>
<td>28</td>
<td>34</td>
<td>22</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>10</td>
<td>22</td>
<td>22</td>
<td>0.00</td>
<td>0.47</td>
<td>29</td>
<td>24</td>
<td>22</td>
<td>0.01</td>
<td>0.37</td>
</tr>
<tr>
<td>11</td>
<td>25</td>
<td>22</td>
<td>0.02</td>
<td>0.32</td>
<td>30</td>
<td>13</td>
<td>23</td>
<td>0.00</td>
<td>0.94</td>
</tr>
<tr>
<td>12</td>
<td>52</td>
<td>22</td>
<td>0.06</td>
<td>0</td>
<td>31</td>
<td>28</td>
<td>23</td>
<td>0.02</td>
<td>0.21</td>
</tr>
<tr>
<td>13</td>
<td>31</td>
<td>22</td>
<td>0.03</td>
<td>0.10</td>
<td>32</td>
<td>23</td>
<td>22</td>
<td>0.01</td>
<td>0.40</td>
</tr>
<tr>
<td>14</td>
<td>32</td>
<td>23</td>
<td>0.03</td>
<td>0.11</td>
<td>33</td>
<td>17</td>
<td>23</td>
<td>0.00</td>
<td>0.81</td>
</tr>
<tr>
<td>15</td>
<td>21</td>
<td>23</td>
<td>0.00</td>
<td>0.59</td>
<td>34</td>
<td>27</td>
<td>22</td>
<td>0.02</td>
<td>0.22</td>
</tr>
<tr>
<td>16</td>
<td>42</td>
<td>23</td>
<td>0.05</td>
<td>0.01</td>
<td>35</td>
<td>27</td>
<td>22</td>
<td>0.02</td>
<td>0.22</td>
</tr>
<tr>
<td>17</td>
<td>20</td>
<td>23</td>
<td>0.00</td>
<td>0.65</td>
<td>36</td>
<td>26</td>
<td>21</td>
<td>0.03</td>
<td>0.19</td>
</tr>
<tr>
<td>18</td>
<td>30</td>
<td>23</td>
<td>0.03</td>
<td>0.14</td>
<td>37</td>
<td>64</td>
<td>22</td>
<td>0.07</td>
<td>0.00</td>
</tr>
<tr>
<td>19</td>
<td>29</td>
<td>23</td>
<td>0.03</td>
<td>0.18</td>
<td>38</td>
<td>57</td>
<td>22</td>
<td>0.06</td>
<td>0.00</td>
</tr>
</tbody>
</table>

There is a pattern here. There are five pairs of coupled questions on the QMCA where the second question asks for the best explanation for the response to the first. These question pairs are 4/5, 13/14, 20/21, 25/26 and 37/38. We see that four of these five question pairs are the items that don’t properly load into a single factor. Question pair 13/14 does not fit into this pattern because question 13 has 5 possible answer choices while the rest have only 2 or 3. The only other question that isn’t adequately described by a single factor is item 12 which asks about how the real and imaginary parts of the ground state of the infinite square well change in time. We theorize that this question isn’t adequately described by the assumed factor because it is primarily assessing understanding of how complex exponentials decompose into real and imaginary parts which is more mathematical understanding than quantum mechanical.

When we increase the dimensionality of the EFA to 2, we find that the second factor loads almost entirely into items 4 and 5 with factor F2 loadings of 0.96 and 0.91 respectively. This means that the additional factor is describing most of the coupled variance of questions 4 and 5 and nothing else. When we increase the dimensionality of the EFA to 3, we find similar behavior where the new factor describes the coupled variance of items 13 and 14. These items have a factor F3 loadings of 0.93 and 0.99 respectively. It is clear from this analysis that the question pairs 4/5, 13/14, 20/21, 25/26, and 37/38 as well as item 12 are not well described by a single factor. In the following Rasch analysis we will first consider the entire set to see if poor fit statistics confirm our results from the EFA, and then we will remove items to see how the remaining set performs under the unidimensional Rasch model.

### C. Rasch analysis

We used the mirt package to do a unidimensional Rasch analysis of our data, generating item difficulties and person abilities based on the fit to the Rasch model. The statistics that we generate for the whole assessment are the M2 statistic which is essentially a modified chi-squared goodness-of-fit test [18], the root mean square error of approximation (RMSEA), and the comparative fit index (CFI) [19]. In general, we want the M2 divided by the degrees of freedom df to be approximately equal to 1, and for relatively good model-data fit, we want an RMSEA < 0.06 and a CFI > 0.9 [19]. When we run the Rasch analysis on the whole test we get M2/df = 4.5, an RMSEA of 0.093, and a CFI of 0.68. All of these indicate poor model-data fit and are consistent with our EFA.

From Table III we have the item fit statistics for the Rasch model. We see from Table III that there are many items that have poor model-data fit because their RMSEA are greater than 0.06. These items are colored red in the table.

Figure 1 shows the estimated item difficulties and person abilities for the whole QMCA under the Rasch model. We see that the person abilities are approximately normally distributed around zero ability, and the item difficulty average is -0.29.

Rather than go through all possible iterations of removing items, we will just discuss removing the fewest number of items in order to achieve the overall model fit statistics of RMSEA < 0.06 and CFI > 0.9. In order to achieve these metrics it was necessary to remove around 14 items. All of
FIG. 1. Estimated item difficulties and person abilities for the whole QMCA under the Rasch model. Note that ability and difficulty are measured on the same scale and presented as such for ease of comparison.

the question pairs 4/5, 13/14, 20/21, 25/26, and 37/38 were removed, and the questions with very high or very low item discrimination as determined by CTT were removed. Those items with high discrimination (ability to distinguish between high and low performing students) were 1, 12, 23, and 27. After removing these items we achieved $M^2/df = 2.3$, an RMSEA of 0.056, and a CFI of 0.902. Table IV shows the individual item fit statistics when these items are removed. We see from this table that all of the items have an RMSEA at or lower than 0.06, which means there is adequate or good model-data fit. Note that the removed items tended to involve time evolution, but not all time evolution items were removed. So we still retained the bulk of content areas in this modified QMCA.

V. CONCLUSIONS

From this analysis we can conclude that 24 of the 38 items on the QMCA are sufficiently unidimensional to be described by a single latent trait under the Rasch model, making them good candidates for a unidimensional test bank. The justifications for removing the poorly fitting items are that they either violate the requirement for local independence, they don’t assess the same dimension as the rest of the questions, or they have extreme discriminations as measured by CTT fit. There are still many analyses that can be done on this data set, including examining whether the subset of people based on their university affects the results, whether the isomorphic spin and wave function context items have similar characteristics, and whether a multidimensional Rasch model would provide better results.

The remaining questions in the QMCA are good candidates for adding to a quantum mechanics test bank that instructors could pull from in order to design their own assessment. Future work for this project includes creating surveys containing all the questions on existing QM assessments so that we can collect data from multiple institutions and perform a similar Rasch analysis on them. We want to have as many compatible questions as possible so that we can create a modular quantum mechanics assessment capable of assessing person ability in a rigorous population independent manner. We also want to generate new items pertaining to second semester quantum mechanics that we can add to this item bank. This will include generating questions, doing student interviews to assess construct validity, and administering to large samples to determine item difficulties.

VI. ACKNOWLEDGMENTS

This work was supported by funding from the Center for STEM Learning and the Department of Physics at University of Colorado Boulder, and the National Science Foundation DUE Grant No. 2143976.

<table>
<thead>
<tr>
<th>Item S-X2</th>
<th>df</th>
<th>RMSEA</th>
<th>p</th>
<th>Item S-X2</th>
<th>df</th>
<th>RMSEA</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>34</td>
<td>15</td>
<td>0.06</td>
<td>0.003</td>
<td>19</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>16</td>
<td>0.04</td>
<td>0.06</td>
<td>22</td>
<td>29</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>16</td>
<td>0.00</td>
<td>0.50</td>
<td>24</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>7</td>
<td>18</td>
<td>16</td>
<td>0.02</td>
<td>0.32</td>
<td>28</td>
<td>31</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
<td>13</td>
<td>0.03</td>
<td>0.19</td>
<td>29</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>15</td>
<td>0.00</td>
<td>0.85</td>
<td>30</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>19</td>
<td>15</td>
<td>0.03</td>
<td>0.22</td>
<td>31</td>
<td>33</td>
<td>16</td>
</tr>
<tr>
<td>11</td>
<td>18</td>
<td>16</td>
<td>0.02</td>
<td>0.32</td>
<td>32</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>27</td>
<td>16</td>
<td>0.04</td>
<td>0.04</td>
<td>33</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>16</td>
<td>38</td>
<td>16</td>
<td>0.06</td>
<td>0.00</td>
<td>34</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>17</td>
<td>26</td>
<td>14</td>
<td>0.05</td>
<td>0.02</td>
<td>35</td>
<td>22</td>
<td>15</td>
</tr>
<tr>
<td>18</td>
<td>9</td>
<td>15</td>
<td>0.00</td>
<td>0.86</td>
<td>36</td>
<td>12</td>
<td>15</td>
</tr>
</tbody>
</table>

TABLE IV. Item fit statistics for the QMCA with items 1, 4, 5, 12, 13, 14, 20, 21, 23, 25, 26, 27, 37, and 38 removed under the Rasch model. S-X2 is the signed chi-squared statistic, df is the degrees of freedom which is variable depending on the binning for S-X2, RMSEA is the root mean square error of approximation, and p is the probability of those data or more extreme values.


Metacognitive knowledge and regulation of peer coaches

Jonathan Kustina and Dawn Meredith

Department of Physics & Astronomy, University of New Hampshire, Durham, NH 03824

Peer coaches are undergraduate peer educators who help facilitate learning in introductory STEM classes, either as learning assistants or peer-led team learning leaders. Peer coaches’ facilitation is generally focused on specific content knowledge, but their pedagogical skills could be applied to other content, such as metacognition. Metacognition, an individual’s awareness and management of their own thinking and reasoning, is an important skill for undergraduate students to learn, though these practices rarely receive the explicit focus required for their development. Peer coaches could act as facilitators of metacognitive practices with their introductory STEM students. As a first step to investigating this potential role, we collected and analyzed written artifacts from the peer coaches’ pedagogical training course, looking for evidence of metacognitive competence. We found that coaches had competence in metacognition both as a learner and as a coach, and that these two perspectives informed each other in productive ways.
I. INTRODUCTION

The ever-present call for better learning environments and educational gains for undergraduate STEM students has led to the development of different types of undergraduate facilitated peer learning. Peer coach facilitators are advanced undergraduates who support active learning in introductory courses. There are several kinds of peer coaches, including learning assistants (LA’s) who facilitate learning during class time alongside professors [1], and peer-led team learning (PLTL) leaders who facilitate learning in small-group sessions outside of normal class hours [2]. These two models have similar goals and outcomes. Peer coaches improve student learning outcomes and interest in their specific topics [1–4]. Other benefits to peer coaching include stronger scientific subject identity in their students and greater retention in science and education majors [1–6]. Peer coaches receive training from a semester-long pedagogy course and weekly preparation meetings with faculty members and senior leaders, ensuring they are prepared to support their students’ learning, problem-solving, and group work.

Learning about metacognition is a recommended part of peer coach pedagogical training [7, 8]. Metacognition describes knowledge and awareness of one’s own cognitive processes [9], a definition that has expanded in recent literature to also include knowledge of specific problem-solving practices and the ability to discern their proper use [10]. Implementing metacognitive practices has been shown to help students by developing higher-order thinking skills, scaffolding of self-regulation, and bolstering motivation for future learning [11, 12], which improves final grades for students [13].

Despite the learning benefits of metacognition, and in contrast to the peer coach pedagogy class, explicit instruction of its practices to students in STEM classroom settings is rare [14, 15]. In the effort to improve student metacognitive practice, we investigated if instructors could productively enlist the help of peer coaches. That is, can peer coaches be effective metacognitive coaches?

Previous studies have investigated similar questions. Lutz and Rios [16] explored LAs’ epistemological growth toward seeing knowledge as co-constructed, which is a foundation for productive metacognitive practices. Another study investigated student metacognitive outcomes when peer tutors were enlisted to support metacognitive growth [17]. Our work will focus on the peer coaches’ preparedness to act as metacognitive coaches, which has not yet been explored.

II. THEORETICAL FRAMEWORK

The framework of STEM teacher knowledge guides our thinking about what it would mean for peer coaches to be metacognitive coaches. This framework posits three facets of teaching knowledge: pedagogical knowledge, content knowledge, and pedagogical content knowledge (PCK) [18, 19]. The latter is an instructor’s specialized knowledge of both content and educational strategies that work best for teaching their specific subject. Previous studies on LAs [20] and PLTL leaders [21] show evidence that peer coaches develop general pedagogical knowledge as part of their facilitation experience. This study focuses on peer coaches’ content knowledge, where the specific content is metacognition. Peer coaches’ PCK about metacognition will be left to future work.

Content competence related to metacognition is conceptualized as composed of two subcategories: metacognitive knowledge and metacognitive regulation [10]. Metacognitive knowledge consists of 1) declarative knowledge about the concrete concepts and practices one knows, 2) conditional knowledge about when and why specific practices should be used, and 3) procedural knowledge about how to implement and manage these practices [10]. Metacognitive regulation consists of 1) planning done prior to cognition in the selection of practices, 2) monitoring one’s awareness and comprehension during a cognitive task’s completion, and 3) evaluating one’s processes and products after cognition for potential improvements [10]. Peer coaches can engage in all of these facets in their roles as learners and coaches. We define metacognitive competence as having both metacognitive knowledge and metacognitive regulation.

Our research questions for this preliminary study are: 1) What metacognitive knowledge and regulation do peer coaches exhibit in their role as learners? 2) What metacognitive knowledge and regulation do peer coaches exhibit in their role as coaches? and 3) How does metacognitive competence in one role inform competence in the other role?

III. METHODOLOGY AND DATA COLLECTION

This study took place at the University of New Hampshire, where the peer coach program serves the introductory STEM courses, with LAs working in physics and mathematics and PLTL leaders working in biology, chemistry, and neuroscience. For their first semester in the program, all peer coaches take the same one-credit pedagogical training course, which focuses on the pedagogical strategies and metacognition they will practice in their new role. The first author, a graduate student of the institution, helped modify and deploy the course and associated materials. The second author is one of the two professors who taught this course over the last 15 years.

In the course there were two class meetings focused on metacognition. The first set of readings [23–25] describe in detail specific implementations of metacognitive practices. Students read one of these papers and then shared their insights through a jigsaw [26] activity with their peers. The second week’s reading was a compilation of cognitive science research on ten easily implemented metacognitive strategies [22]. In class, the peer coaches discussed each strategy, including when and why it should work.

We collected all written assignments from the pedagogy course from those peer coaches who consented to partici-
TABLE I. Description of written artifacts collected from peer coaches’ pedagogical training.

<table>
<thead>
<tr>
<th>Data Name</th>
<th>Prompt</th>
</tr>
</thead>
</table>
| Synthesis Paper               | 1) A status report on your progress as a facilitative leader. Speak particularly to how you may have change or evolved in your actions or thinking. In particular, I want you to reach to incorporate ideas from any theories or research on student learning that we read about or discussed.  
2) How participating in this course and having this experience has affected your own learning in other courses. This is about YOU as a student, NOT in this pedagogy course |
| Reflection on Metacognitive  | Q1: Which reading did you do?  
Q2: What was something new in this reading for you, and why did you find this interesting or important?  
Q3: Give details about how some part of this reading connects to your previous beliefs and/or experiences.  
Q4: Ask an "I wonder" question. |
| Readings                      | Q1: What is one study strategy that works well for you and how/why does it help you?  
Q2: What is one study strategy that hasn’t worked for you and how/why didn’t it help?  
Q3: How do you know when you know something? For example, when do you feel confident that you understand a metabolic pathway, or a chemical reaction, or a mathematical proof, or a physics problem solving strategy? |
| Self-report of Effective Study Strategies | Q1: What is one study strategy that works well for you and how/why does it help you?  
Q2: What is one study strategy that hasn’t worked for you and how/why didn’t it help?  
Q3: How do you know when you know something? For example, when do you feel confident that you understand a metabolic pathway, or a chemical reaction, or a mathematical proof, or a physics problem solving strategy? |

IV. ANALYSIS

Through our analysis, we categorized instances of peer coaches’ metacognitive competence around five key features. These are metacognitive knowledge and metacognitive regulation, each from both the perspective of a learner and the perspective of a coach, along with the interplay between these two perspectives.

A. Learner Metacognitive Knowledge

We define learner metacognitive knowledge as a peer coach’s knowledge of learning strategies that utilize higher-order thinking skills, including how to use these strategies themselves, when to use them optimally in their learning, and why they are effective for their learning. Peer coaches’ learner metacognitive knowledge can come from their own experience as learners and their pedagogical training. We have seen evidence of this throughout the written artifacts collected, with both Ava and Stephanie making mention of multiple study strategies they utilize when studying for their own classes:

- Group Work
- Practice Testing
- Flashcards
- Vocalization
- Self-Explanation (how is new information related to old information, explaining steps in problem-solving)
- Summarization
- Rereading/rewriting notes
- Chunking (breaking down difficult tasks to smaller tasks)
- Imagery to Text

These strategies are detailed in two of their readings for the pedagogy course [22, 25]. The two peer coaches demon-
strate their knowledge of many strategies, going significantly beyond the common strategies of re-reading and highlighting [22], suggesting a strong foundation for metacognitive competence.

The peer coaches show further learner metacognitive knowledge in their understanding of the use and effectiveness of these study strategies. For example, Stephanie responded to a metacognitive reading [22] with the following:

Something new I learned from this reading was the process of learning and how re-reading something over and over again is not the correct way of learning. I found this very important because that’s how I’ve been studying for my exams by cramming in the material by rereading lecture slides in which did not help me in the long run.

Stephanie’s learner metacognitive knowledge allows her to recognize strategies like rereading and cramming, as well as judge them for her own learning and adjust accordingly. She demonstrates the active processes of utilizing content knowledge of metacognitive regulation, which is indicative of metacognitive competence.

B. Coach Metacognitive Knowledge

We define coach metacognitive knowledge as a peer coach’s knowledge of classroom facilitation strategies and pedagogy, which includes what strategies are effective for a specific topic, when students can best utilize them, why such strategies are effective in different contexts, and the peer coach’s individual philosophies about facilitation. Both Ava and Stephanie wrote about metacognitive practices they have seen within the context of coaching, including:

- Group Work
- Practice Problems
- Elaborative Interrogation (explaining why something is true)
- Chunking
- Visualization

This list shows the familiarity that peer coaches have with metacognitive practices that occurred during coaching. These discussed strategies were either observed in students’ behavior or introduced by the peer coaches. Their discussion of these strategies demonstrates their ability to recognize practices regardless of origin, and therefore demonstrates their coach metacognitive knowledge.

We see additional evidence of coach metacognitive knowledge as our peer coaches discuss the implementation of these known strategies in their coaching sessions. For example, when speaking on the decision to include more visual components to her problems, Stephanie writes:

[H]aving something visual to plan out and followed really helped my students who like to have something to look at while solving a problem and if they did encountered something similar on a homework problem or a question on an exam they can visualize back to our activity and remember the steps on how to tackle it.

Here, Stephanie shows her coach metacognitive knowledge by drawing a connection between the learning strategies she can facilitate in her sessions and their benefits to her students. Peer coaches demonstrate metacognitive competence through not just knowledge of metacognitive practices, but of the pedagogical implications during the coaching experience.

C. Learner Metacognitive Regulation

Learner metacognitive regulation is a peer coach’s active thoughts and behaviors relative to learning and performance on cognitive tasks. This includes planning with consideration for familiarity and interest in the topic, monitoring for actions and behaviors that promote or deter learning, and evaluating work from the context of a course and one’s goals. Learner metacognitive regulation can be seen when peer coaches reflect on their active learning behaviors over a semester. For example, when Ava talks about her use of practice problems to study for her introductory physics course, she says,

Not only was it easier to memorise the formulae and apply them, but also the theory started to make more sense. I consciously understood the best study strategy for me and worked on areas that I found difficult.

Here, Ava engages in learner metacognitive regulation by actively monitoring her progress in a new study strategy, then evaluating its effectiveness so that it can be better tuned to her needs as a learner.

D. Coach Metacognitive Regulation

Our definition of coach metacognitive regulation is a peer coach’s active facilitation strategies and behaviors. This includes planning instruction around both content and pedagogy, monitoring their facilitation’s effectiveness through the responses from student learning, and evaluating the impact of their facilitation strategies based both on student assessment and personal reflection. We found evidence of coach metacognitive regulation in the peer coaches’ work with their students, such as when Stephanie responded to a reading assignment focused on cooperative learning:

Group effort, there is a saying that a group is as strong as the weakest link. I’m not saying that a person is dragging the group down but they should help a group member who is having a hard time with the topic or subject and from that rewards the student with extra credit and from there reflects their own individual goals and I see
The benefits of group work, Ava wrote, impacted their own learning. For example, when speaking on how their implementation of coaching strategies impacted the value of the strategies for her own learning, she was more likely to extend the value of those strategies to her students. A peer coach’s ability to metacognitively reflect on their own learning can provide great insight into the practices that might benefit their students, enabling a stronger coaching experience.

V. CONCLUSIONS AND DISCUSSION

This pilot study shows that our two peer coaches demonstrate several aspects of metacognitive competence. This includes metacognitive knowledge and metacognitive regulation, both from the perspective as a learner and as a coach, along with the interplay of these two roles. Evidence of these facets of metacognitive competence was seen in the written artifacts from their pedagogical training course, where they spoke on their experiences from both roles over their first semester.

We have seen similar evidence in other peer coach reflections as well. In future work, we will analyze the remainder of the data set (N=127) to look for further evidence of metacognitive competence. We will also focus on the uniquely powerful interplay of perspectives, and seek evidence that this interplay allows for peer coaches to experience increased metacognitive growth compared to an individual holding only one such role.

Building from this work and the literature, we will go beyond the current research to construct an fuller operational definition of a metacognitive coach (which will include metacognitive competence and interplay as core components of facilitating metacognition) and analyze the data from the full set of students to investigate to what extent our peer coaches are metacognitive coaches. Other characteristics which we will investigate include the peer coaches’ ability to be explicit about their metacognitive competence and to tie the formal readings to their own experiences; both of these characteristics indicate deep understanding [27, 28]. We have also seen hints of pedagogical content knowledge [19] about teaching metacognition. For example, some peer coaches show an awareness of unproductive but common metacognitive practices, and others show a belief that metacognition should be explicitly taught.

The end goal of our work is to inform classroom practices both in the pedagogy course and in the peer coach interactions with students. Future work could investigate which readings and reflection prompts in the pedagogy course are effective for developing productive coach practices and peer coaches’ deployment of metacognitive interventions with their students.

ACKNOWLEDGMENTS

This work was supported by NSF Grant 2013427. The authors gratefully acknowledge conversations with Amanda Gaudreault, and Drs. Kathleen Bowe and Brock Couch.
[26] https://jigsaw.org/#steps Retrieved 05/18/23
Physics Teachers’ Motivations to Learn Computational Thinking as a Re-novicing Experience

W. Brian Lane\textsuperscript{1,2}, Terrie M. Galanti\textsuperscript{2,3}, and XL Rozas\textsuperscript{2}
\textsuperscript{1}Department of Physics
\textsuperscript{2}Northeast Florida Center for STEM Education
\textsuperscript{3}Department of Teaching, Learning, & Curriculum
University of North Florida,
1 UNF Drive, Jacksonville, FL, 32224

Integrating computational thinking (CT) into high school physics requires physics teachers to develop specialized pedagogical content knowledge of computing and content integration. We describe this as a re-novicing experience, in which teachers build and act upon knowledge at the intersection of physics, computation, and pedagogy. We observed this re-novicing in a Python workshop for physics teachers (n = 23) from three countries. Teachers participated in computational learning activities in Jupyter Notebooks with the goal of developing their capacity to integrate Python in physics applications. In this exploratory study, we describe a framework for integrated CT professional learning and report teacher responses to a post-workshop survey about their motivations for participating in this re-novicing process. Their motivations included needing to learn essential programming features and an intrinsic motivation to transform how they teach physics. These findings suggest that professional learning for CT integration needs to interweave computing content with pedagogical applications.
I. INTRODUCTION AND FRAMEWORK

Computational Thinking (CT) is the processes of using logic and algorithms to formulate problems and construct solutions that a computer can carry out [1]. In physics practice and education, computer programming is the vehicle by which such solutions are enacted. The integration of CT [2–9] into undergraduate physics education is an established practice [10–13]. However, because this integration is a relatively recent development [14] and has yet to fully reinforce CT as relevant for future educators [15], the integration of CT into the high school physics context is still in the beginning stages [8, 16–20]. As CT integration is not required by any district or curriculum found in the United States, integrating CT is an instructional choice that requires teachers to learn computer programming after graduating.

The term re-novicing has been used to describe when teachers are assigned to subjects or grade levels for which they lack disciplinary content knowledge or discipline-specific pedagogies [21–23]. A teacher’s decision to learn how to integrate computer programming within physics instruction is a form of re-novicing.

The re-novicing process we discuss in this paper begins with learning computing practices and proceeds toward full physics-CT integration [24]. This process is understandably challenging, and supporting teachers requires understanding the factors that might motivate them to begin learning about CT integration and sustain their implementation of it. In this paper, we ask the research question, What motivates physics teachers to learn computing as a re-novicing experience?

To understand teacher motivations, we consider three domains of learning required to integrate CT with physics as a disciplinary subject. Grover [25] frames the integration of CT with a disciplinary subject (like physics) by extending definitions of technological pedagogical content knowledge (PCK) [26] to include the subject domain and computing as distinct but overlapping fields. As depicted in Figure 1, this CTIntegration framework depicts three categories of PCK: (1) Physics Content & Practices (such as one might learn in an undergraduate physics program), (2) Computing Content & Practices (such as one might learn in an undergraduate computing program), and (3) Pedagogy (such as one might learn in an undergraduate education program). The intersections in this framework highlight the interdisciplinary knowledge bases of Physics-Computing Integration (1 ∩ 2, how to solve a physics problem using iterative numerical calculations), Computing PCK (2 ∩ 3, how to teach programming practices), Physics PCK (1 ∩ 3, how to teach physics), and Physics-CT Integration PCK (1 ∩ 2 ∩ 3, how to teach physics and computation in an integrated manner that enables new avenues of learning [27]).

In our prior work [24], we theorized a pathway of re-novicing (dashed arrow in Figure 1) to describe individual teachers’ professional learning needs through these domains. This pathway begins by introducing computing content and practices (Circle 2 in Figure 1), experiencing physics applications of computing (Intersection 1 ∩ 2), and developing the PCK necessary to deliver physics-CT integrated learning strategies in their courses (Intersection 1 ∩ 2 ∩ 3). For example, suppose a high school physics teacher wanted to integrate CT into their unit on projectile motion. They must first learn the necessary computing practice of iterative loops (Circle 2), then apply an iterative loop to automate a numerical solution of the projectile’s motion (Intersection 2 ∩ 3). Finally, they must reflect on how they can teach students to use this approach to understand projectile motion (Intersection 1 ∩ 2 ∩ 3). This re-novicing pathway also helps us categorize teachers’ motivations to integrate CT in physics and thereby provide appropriate professional learning opportunities.

II. RESEARCH CONTEXT

We conducted an online synchronous workshop for high school physics teachers in summer 2022, co-sponsored by the American Association of Physics Teachers [24]. The physics teachers enrolled in the workshop (n = 23) were from 15 different states in the United States and from Canada and Italy. All participants were experienced in-service high school teachers and AAPT members. Their self-reported experiences with computer programming prior to the workshop varied from no experience to formal computing coursework. We structured each day of this three-day workshop to feature

![FIG. 1. Modified CTIntegration framework [24, 25] to depict a pathway of re-novicing (dashed arrow) to describe individual teachers’ professional learning needs through these domains. This pathway begins by introducing computing content and practices, experiencing physics applications of computing, and developing the PCK necessary to deliver physics-CT integrated learning strategies in their courses.](image-url)
a three-hour morning session and a three-hour afternoon session. During morning sessions, teachers worked in groups of 4-6 in breakout rooms to complete a computational activity designed to be accessible in a high school physics environment. These activities were based in Jupyter notebooks so that teachers could experience the integrated physics/CT curriculum as learners [28] in a collaborative online setting. During afternoon sessions, teachers were offered a choice board of computational integration topics, including assessment of computation-based activities, lab adaptations, visualizations of electrostatics, random number generation, quantum applications, and the use of external data sets. These sessions were designed to be exploratory, with each group working at their own pace and relying on each other’s insights and a facilitator’s feedback.

The learning activities were designed to build teachers’ capacity [29] for CT integration in physics-situated computing applications, and the workshop structure was designed to model computing PCK [30]. This capacity development is a re-novicing experience [21], in which teachers who aspire to integrate new content may struggle in similar ways to new teachers as they negotiate gaps between what they already know and what they have yet to learn. Professional learning that supports teachers on this pathway of re-novicing requires an understanding of their motivations to integrate CT.

III. METHODOLOGY

We collected survey responses from \( n = 23 \) teachers who participated in a three-day workshop about physics applications of programming in Python. In this paper, we focus on their responses to the first survey question, “What first motivated you to participate in the workshop?” using qualitative content analysis. The first author used first cycle open concept coding [31] to identify common themes among responses. This author relied on his observations as a workshop leader to contextualize the teachers’ comments. The three authors discussed these open codes and agreed that they adequately characterized the teachers’ responses. We then used the domains of our adapted CTIntegration framework [24] as a set of second cycle provisional codes [31] to map teacher motivation onto professional learning needs. The first two authors independently coded the first cycle concept codes using the circles and intersections (see Figure 1) by re-examining the associated teacher responses. To increase the credibility and dependability [32] of our qualitative findings, the third author provided a peer debriefing [33] of the data analysis using his expertise as a STEM education researcher. The first and second author responded to the third author’s questions and revised coding accordingly until consensus was reached.

IV. RESULTS: TEACHERS’ DESCRIPTIONS OF MOTIVATION

Here, we discuss themes in teachers’ responses to a post-workshop survey question: “What first motivated you to participate in the workshop?”

A. First-Cycle Thematic Coding

In our first cycle coding, we identified seven conceptual themes (bold text below and first column of Table I) among the teachers’ responses. Some responses exhibit multiple themes, so the sum of the numerical values reported below exceed our \( n = 23 \) participants.

Four teachers expressed an interest in beginning to learn programming, which indicated that the teacher wanted to learn programming for the first time. These teachers felt they were complete novices with respect to computing content and practices. Four other teachers explained that they wanted to build on their prior programming experiences as learners. These responses indicated prior experience with programming but an interest in refreshing or expanding this knowledge. This finding supports prior research that reveals how experiencing CT as an undergraduate physics student may be insufficient for a physics teacher to integrate CT into their teaching [15], and how physics teachers learn to integrate CT by exploring their own interests and capabilities first [20].

The greatest number of teachers (13 out of 24) described wanting to teach physics differently thanks to the affordances of computational approaches. Some teachers described this as a long-term goal, or as an opportunity created by curricular changes, or a desire to improve the quality of teaching and learning. Three teachers expressed an interest in building on prior programming experience as teachers. These responses indicated either prior experience with integrating CT into a physics class or teaching a computing class.

Three teachers expressed an interest in teaching a physics course with computation. These teachers specifically identified CT integration as a central component of an upcoming course, indicating they were already committed to implementing what they learned in the workshop.

Seven teachers described the importance of preparing students for postsecondary STEM courses or professions. These responses highlighted the importance of computing applications in the their students’ future education and careers. Two teachers specifically mentioned responding to student interest after students or alumni requested programming-related curricula.

B. Second-Cycle Provisional Coding

In our second cycle provisional coding, we identified where in the CTIntegration framework (Figure 1) each of the
TABLE I. First and second cycle coding of \( n = 23 \) teacher responses to survey question: “What first motivated you to participate in the workshop?” First cycle codes summarize themes in teacher responses. Second cycle codes map these themes onto the CTIntegration framework in Figure 1. Nearly all (20 out of 23) teacher comments relate to the central provisional code of Physics-CT Integration PCK.

<table>
<thead>
<tr>
<th>First Cycle Concept Codes (Number of Participants) with Illustrative Quotes</th>
<th>Second Cycle Consensus Provisional Codes (Number of Participants)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Begin to learn programming</strong> (4)</td>
<td><strong>Circle 2 Computing Content &amp; Practices</strong> (10)</td>
</tr>
<tr>
<td>“I want to familiarize myself with a programming language.”</td>
<td></td>
</tr>
<tr>
<td><strong>Build on prior programming experience as a learner</strong> (4)</td>
<td></td>
</tr>
<tr>
<td>“I decided I needed to get my students into the 2020s and to do so, I had to dust my quarter-century old computing skills off and learn some Python.”</td>
<td></td>
</tr>
<tr>
<td><strong>Respond to student interest</strong> (2)</td>
<td></td>
</tr>
<tr>
<td>“I was approached by a former student who had just obtained his undergrad degree, who remarked, ‘your class was great ... it really would have been nice if we had some coding in it as well.’ Honestly, that really lit the fire in me.”</td>
<td></td>
</tr>
<tr>
<td><strong>Build on prior programming experience as a teacher</strong> (3)</td>
<td><strong>Intersection 2 ( \cap 3 ) Computing PCK</strong> (3)</td>
</tr>
<tr>
<td>“I have always dabbled in the use of computational physics in my classes, but never had a clear vision as to what it might look like in regards to content, platform/language, etc.”</td>
<td></td>
</tr>
<tr>
<td><strong>Teach physics differently</strong> (13)</td>
<td><strong>Intersection 1 ( \cap 2 ) \cap 3 Physics-CT Integration PCK</strong> (20)</td>
</tr>
<tr>
<td>“It has been a goal for a few years to include computational problem solving skills in my physics courses but the barriers to entry felt high... I thought this workshop could give me enough tools to get over my initial hesitations to start something, even if it’s not perfect.”</td>
<td></td>
</tr>
<tr>
<td><strong>Prepare students for postsecondary STEM courses/professions</strong> (7)</td>
<td></td>
</tr>
<tr>
<td>“Alumni from my classes regularly come back from university, internships, etc. indicating that programming experience was essential to their success in STEM post-high school. Or they indicate that they wish they had more exposure to programming skills in high school because much of it is assumed knowledge in their college STEM coursework and they have to figure it out on their own. I want to build my Python skills so that I can add programming seamlessly into courses students are already taking at my school (like Physics).”</td>
<td></td>
</tr>
<tr>
<td><strong>Teach a physics course with computation</strong> (3)</td>
<td></td>
</tr>
<tr>
<td>“I have a certificate in computer programming and have always loved it, but had trouble figuring out how to incorporate it.”</td>
<td></td>
</tr>
</tbody>
</table>

First-cycle themes seemed to indicate the teachers’ motivations lie (second column in Table I). We found that three areas (one circle and two intersections) in the framework seemed to describe these themes.

The themes of beginning to learn programming, building on prior programming experience as a learner, and responding to student interest lie within Circle 2, Computing Content & Practices. As described earlier, this category comprises computing knowledge independent of disciplinary applications or pedagogical delivery. These themes indicate a focus on teachers’ need to develop knowledge of programming or students’ expressed need to develop such knowledge. While the teachers attended a workshop designed to deliver and utilize such knowledge in the context of a physics course, these motivations would also be satisfied in a computing science context without physics applications.

The theme of building on prior programming experience as a teacher lies within Intersection 2 \( \cap 3 \), Computing PCK. This theme highlights a few teachers’ experience with teaching computing topics and wanting to expand this capacity. Responses within this intersection focused on the technicalities of computing-based instruction, such as platforms and languages.

Finally, the frequently encountered themes of teaching physics differently, preparing students for postsecondary STEM courses or professions, and teaching a physics course with computation lie within the central Intersection 1 \( \cap 2 \) \cap 3, Physics-CT Integration PCK. These themes indicate a holistic, interdisciplinary view of what CT integration can accomplish in a physics course.

V. DISCUSSION AND CONCLUSIONS

In the themes discussed in Section IV A, we see a combination of intrinsic motivation and extrinsic motivation. For example, nine teachers framed their interest in learning programming as a response to requests for greater exposure to programming in high school from both their current and former students. This identified need for early exposure to programming represents an extrinsic motivation based on factors outside of the teachers’ classrooms. One teacher shared, “I’ve had a few students who were interested in programming (even participating in competitions), and I wanted to take advantage
of that.” Another teacher focused on former students’ lack of preparation for STEM postsecondary education: “The number of former students that have gone into computer-related majors or computer-related careers is high and yet we offer almost no computer coding classes at any of our district’s comprehensive campuses.” This extrinsic motivation indicates that teachers want to make their high school physics experience relevant for students’ future learning needs.

Evidence of extrinsic motivation was not limited to students’ expressions of interest. Other teachers were similarly motivated by recommendations from STEM professionals. One teacher recalled viewing a webinar from the Department of Energy which encouraged teachers to integrate programming into what they were already doing in the classroom. Another teacher commented, “Every time I have researchers come in to talk with my students, they all say how useful it is to know some coding and that Python is used by a lot of labs these days.”

Although extrinsic motivation related to student interest and professional recommendation was expressed by some teachers, over half (13 out of 24) of teacher responses speak to an intrinsic desire to transform physics education using CT. These teachers reported seeing the physics-relevant benefits of CT as a reason for attending the workshop. For example, four teachers described wanting to integrate CT into their AP Physics courses to improve the quality of their teaching and learning, with one teacher calling this a “re-imagining” of their AP program. Another teacher had been looking for a way to integrate CT into AP Physics for over a decade. Their perception of a gap between students with and without coding experiences had prevented this teacher from integrating CT. They explained how Jupyter notebooks, as a web-based application, would support more equitable access to programming for students with a wide-range of prior experiences: “Jupyter notebooks seem like a way around this - browser-based, prewritten code that could be altered, with more efficient scaffolding.” Jupyter notebooks were critical to our workshop’s focus on Intersection 1 ∩ 2 ∩ 3, Physics-CT Integration PCK. Our workshop approach was responsive to teachers’ motivations to teach physics differently and provide authentic programming experiences for students.

Overall, these teachers entered the workshop already aware of what CT integration could accomplish in their courses and they were eager to realize those benefits. In total, 20 of the 23 teachers’ responses touched on a theme related to Intersection 1 ∩ 2 ∩ 3 in Figure 1. Even before the workshop began, their motivations aligned with our workshop philosophy. They recognized the benefits of CT integration even while 10 of the 20 teachers described wanting to learn foundational computing content and practices (Circle 2). This contrast further illustrates our pathway (dashed arrow in Figure 1), on which they can clearly see the end goal of Intersection 1 ∩ 2 ∩ 3 even while their professional learning begins at Circle 2 [24]. This motivation is critically important for teachers to persist in the process of learning computing concepts and carrying out CT integration in their courses throughout the school year.

We also note the regions of the CTIntegration framework that the teachers’ responses did not address. They did not, for example, report interest in topics at Intersection 1 ∩ 2, physics-computing integration. Although this overlap of physics knowledge and CT knowledge is important in the process of integrating CT into a physics course, the teachers seemed more interested in getting to the pedagogical use of such applications, rather than the applications themselves. It seems that these teachers wanted to learn programming in the context of teaching physics. Most did not mention needing to learn programming, but already envisioned the integration and were less worried about how much Python they needed to learn.

We acknowledge that the teachers in this study self-selected into a summer computer programming workshop and that these results may not be generalizable to all high school physics teachers. Even though these teachers were already interested in integrating CT in physics curriculum and instruction, these observations have important implications for offering professional learning. It seems that, for many teachers, professional learning needs to interweave computing content and practices with pedagogical applications. It is not necessary, for example, to offer every teacher an introduction to a programming language devoid of applications, nor for each teacher to complete a computational physics research project. The pathway of re-novicing that we have established explains both the teachers’ learning process [24] and their motivations for beginning this process.

We have described an online workshop for high school teachers to make progress along a pathway within a CTIntegration framework. This pathway is a re-novicing experience that begins with requisite computing knowledge, progresses through physics applications of computing knowledge, and arrives at physics-CT integration pedagogical content knowledge. We collected teachers’ self-reported motivations for participating in this workshop and used first cycle open concept coding to generate themes in these responses. We then used second cycle provisional coding to map these themes onto teacher professional learning needs. We found that teachers’ self-reported motivations for participating in this workshop focused primarily on the beginning and endpoints of our outlined pathway, with a combination of extrinsic and intrinsic motivational factors. These motivational factors are important considerations in professional learning design and implementation of CT-physics integration in the high school context. Understanding physics teachers’ motivations to learn computing and begin CT integration supports the design of professional learning tailored to their preferences and experiences.

ACKNOWLEDGMENTS

This work was supported by the Voya Foundation and the American Institute of Physics.
Exploring alternative perspectives through fictionalized student dialogues

Thanh Lê and Andrew Boudreaux
Department of Physics and Astronomy and Science, Math and Technology Education Program, Western Washington University, Bellingham, WA, 98225

Jayson Nissen
Nissen Education Research and Design, Slidell, LA, 70458

Research-based instructional strategies often use fictionalized student dialogues (FSD) to encourage students to explore diverse ideas and perspectives. This study investigates the effectiveness of one such FSD in facilitating the exploration of alternative perspectives on the change in speed of a cart tapped by a finger. In one perspective, change occurs continuously during the tap, while in the other, change occurs instantaneously. We collected classroom video of eight groups discussing the FSD, and apply the socio-metacognitive framework of Borges et al. in our analysis. Five groups exhibit moderate- to high-level exploration of the contrasting perspectives, including one group that challenges the instantaneous perspective by drawing on the real-world experience of driving a car. Conversely, groups with low- to moderate-level exploration discontinued exploration upon seeking the “right” answer. Findings suggest that FSDs can support the exploration of alternative perspectives, and that instructors can enhance exploration by emphasizing real-world experiences and discouraging “answer-making.”
I. INTRODUCTION

Student-centered, research-based materials and curricula often use fictionalized student dialogues (FSDs) to present students with multiple perspectives on a physical phenomenon. Tutorials in Introductory Physics [1], Next Generation Physical Science and Everyday Thinking (NGP) [2], and TIPERs [3] all use FSDs. An FSD encourages students to explore contrasting perspectives and ways of reasoning about a specific concept or phenomenon, a habit of mind that underlies many scientific practices [4]. In an FSD, two or more fictitious students present contrasting explanations, which are then discussed and evaluated by the students. The contrasting perspectives generally derive from the research base: the fictitious students voice specific lines of thinking known through research to be common. FSDs have been useful in part because specific difficulties are often persistent. Research shows that simply pointing out an incorrect line of reasoning, and emphasizing how it differs from the normative reasoning, does not usually improve functional understanding [1, 5]. Once learners adopt a flawed explanation, they can be vulnerable to confirmation bias, in which they focus on supporting the initial explanation and do not actively consider alternatives [6, 7]. While research has shown the efficacy of activities that include FSDs [8], details of the learning dynamics have to our knowledge not been studied.

Socio-metacognition refers to how groups monitor and regulate their interactions and collective learning processes [9–12]. Theories of socio-metacognition align with the sociocultural nature of student-centered instruction, including FSDs. One such framework, from Borge and colleagues, includes a communication pattern of particular relevance to our study: exploring alternative perspectives (EAP). EAP refers to the extent to which a group presents and discusses alternative opinions, claims, and ideas [11].

Groups with a strong tendency toward EAP may be more likely to benefit from learning activities that involve FSDs. To investigate the role of exploring alternative perspectives in groups’ engagement with fictionalized student dialogues, we asked the following research question: To what extent does a fictionalized student dialogue, with two contrasting ideas about the concept of change in speed, lead students to explore alternative perspectives? Understanding the details of how FSD activities do or do not promote the exploration of alternative perspectives can support instructors in effective facilitation of materials that use FSDs, and inform the design of curricula that elicit higher levels of EAP.

II. THEORETICAL FRAMEWORK: SOCIO-METACOGNITION

We use the socio-metacognitive framework proposed by Borge and colleagues [11, 12]. At an individual level, metacognition involves a learner’s knowledge about cognition, as well as the skills, or executive processes, that the learner deploys to monitor, control, and regulate cognition [13–15]. During collaborative learning, group members externalize their thinking through discourse and actions. The group either engages with or dismisses these externalized ideas via social interactions [12]. How the group monitors and regulates their interactions and collective learning processes is referred to as socio-metacognition. In this framework, the production of collective knowledge requires negotiation, which impacts how participants involve themselves in the collaboration. Communication patterns activate specific forms of collaboration and cognition. The framework focuses on two communication macro-patterns: collective information synthesis and collective knowledge negotiation. Each consists of three micro-patterns. Collective information synthesis occurs when members of a group share individual information, check their understanding and synthesize the information to develop shared understanding and new knowledge. Collective knowledge negotiation occurs when a group explores alternative perspectives, proposes and evaluates ideas, and engages in constructive discourse. Because FSDs propose two (or more) alternative explanations for a physical phenomenon, they (ideally) lead students to compare, analyze, and critique the alternatives. Thus, our study focuses on exploring alternative perspectives (EAP): the extent to which a group presents and discusses alternative opinions, claims, and ideas.

III. METHODS

Next Generation Physical Science and Everyday Thinking (NGP) is a research-based, student-centered curriculum, intended for pre-service K-8 teachers and consistent with the Next Generation Science Standards [16]. Core NGP activities focus on physics and physical science content, and supplementary activities explore the nature and science of learning. NGP adopts a social-constructivist model: students conduct guided experiments using simple equipment and computer simulations in small groups, punctuated by full-class discussions where groups present their findings and seek consensus. NGP targets concepts and reasoning research has identified as challenging for many physics learners [16], often through the use of fictionalized student dialogues (FSD).

This study involved physics courses using NGP at two institutions on the west coast. Both are regional, public, primarily undergraduate universities with a focus on teacher preparation. In all courses, instructors acted as facilitators, listening in on small-group discussions, responding to student questions, posing additional questions intended as formative assessment or guides to learning, and leading full-class, consensus discussions. At both institutions, the courses included the NGP units on Newton’s laws and energy. This study focuses on the fictionalized student dialogue found in Unit EM: Energy-based Model for Interactions, Activity 1: Interactions and Motion (UEM-A1). UEM-A1 is the entry point for students’ study of energy concepts, and students’ first exposure
FIG. 1. Step 3 from NGP UEM-A1 includes the fictionalized student dialogue of two contrasting ideas about the concept of change in speed.

Consider this conversation between two students, who were discussing their ideas of how the speed of the cart changed when it was given a quick tap in the direction it was already moving, and its speed increased from 20 cm/s to 40 cm/s.

I think the speed changed gradually over a very short period of time. At different points in time during this period the speed of the cart passed through all the values between 20 cm/s and 40 cm/s.

I disagree. Everything happened so quickly that I think the speed changed in an instant. One moment it was 20 cm/s, the next it was 40 cm/s. It was never anything in between these two values.

Kristen

Amara

Sketch speed-time graphs illustrating the difference between Kristen’s and Amara’s ideas. Which of them do you agree with (or do you agree with neither of them) and why?

IV. RESULTS

In this section, we describe three case studies we classified as representing high, low, and moderate-level EAP based on the extent to which they discussed both ideas, and identify themes within each. In the Discussion, we present a comparative analysis. Quotes are lightly edited for readability.

Case Study 1: High-level exploration of alternative perspectives. In this case study, three students, Alma, Barbara, and Cathy, spend just over six minutes discussing the instantaneous change perspective of Amara and the gradual change perspective of Kristen.

The group leads by agreeing with one of the two perspectives. Alma initiates the discussion by endorsing the gradual change perspective (“I think it’s Kristen”). Cathy immediately confirms this by disagreeing with the contrasting, instantaneous change perspective (“it’s not at an instant”). As a group, the students then explore each perspective in detail, first by applying each perspective to a familiar real-world phenomenon (driving a car), and then by drawing speed-time graphs for each perspective.

Alma introduces the real-world context, which is then immediately taken up by both Cathy and Barbara:

Alma: When you are driving a car you’re not going to press on the gas pedal when you are going 28 and [go] immediately to 56 without even going through 29.

Cathy: Is that even possible for any type of circumstance, for it to just be...

Barbara: Not like instantly... even if you are in a fast car it goes gradually.
stantaneous change perspective. Cathy takes up Alma’s real-world example by generalizing it (“Is that even possible for any type of circumstance?”). Barbara then continues by offering validation of the alternative perspective (“Even if you are in a fast car, it [the speed] goes gradually”). In this first part of the case study, which lasts about 80 seconds and involves 28 talk turns that include all members, the group explores the implications of the contrasting perspectives presented in the fictionalized dialogue, rather than only validating the perspective with which they initially expressed agreement. The group compares the implications of Amara’s instantaneous change perspective with their own real-world experience, and the recognition of inconsistency leads them to reject Amara’s perspective.

In the second part of Case Study 1, the group constructs the two speed-time graphs. The group explores both graphs in detail, rather than only discussing the gradual perspective that they agree with. The students start by individually sketching graphs for each perspective, and then compare their graphs to check for agreement. The students discuss details of the shape of each graph, including how underlying perspectives on change relates to the shape of the graph. For example, when discussing the graph Amara would draw, Barbara explains “So one moment it [the speed of the cart] was at 20 and next - so like it would be... there wouldn’t even be a line ‘cuz it’s not gradual, right?”’, to which Alma agrees.

We categorize Case Study 1 as high level EAP due to the sustained consideration of two contrasting perspectives on change in speed. Although the group has already reached a consensus on which perspective makes more sense to them, they explore each perspective in detail.

Case Study 2: Low-level exploration of alternative perspectives. Our low EAP group consists of three students, Kate, Amy, and Jose. They spend three minutes on the FSD, one minute agreeing with Amara’s instantaneous idea and two minutes drawing the speed-time graphs.

The group discussion starts after Amy reads the question aloud, and then immediately shares that she thinks the speed changes at an instant based on her experience pushing the cart on the track (“When we pushed, there was not really change and it was just going straight, probably the only one that was changed was when we put our finger in that there was instant change”). Jose agrees with her, pointing out that this is aligned with Amara’s idea, and Kate adds that the speed change was “very fast.”

This discussion ends when Amy reminds the group to draw the speed-time graph (“Now we have to sketch the speed-time graphs”). After the students have all independently drawn their graphs, Jose checks in with his partners. While holding his paper up to show his graphs, he explains how he drew them based on the timescale. He says, “I think Kristen’s might take longer, so it’d be more like that, a little flatter, and then Amara, since it’s at an instant, like how we drew ours, it’d be more like right away because it takes less time, and that takes longer ... because it took less time for Amara and that’s what she thinks and then it took more time for Kristen’s to gain speed.” Amy and Kate agree with Jose’s graphs and explanations (“I think that makes sense”).

We categorized Case Study 2 as low EAP due to the lack of interrogation of the gradual change idea. After endorsing Amara’s instantaneous change idea, the group does not explore Kristen’s alternative perspective of gradual change. Although the students agree the change in speed is instantaneous, Kate seems to raise the possibility of Kristen’s gradual change idea by describing the change as “very fast.” However, the group does not discuss what the gradual change would look like or why a gradual change might be correct. The prompt in the activity seems to force the students to consider both perspectives, and they describe differences between the two graphs. However, the students do not interrogate Kristen’s idea of gradual change in speed, or discuss in detail how or why that idea does or does not make sense to them. At no point does the group engage with the abstract notion of an instant in time, or consider differences that arise from analyzing the motion at shorter or longer timescales.

Case Study 3: Moderate-level exploration of alternative perspectives. This group of four students, Cami, Penny, Huy, and Brian, explores alternative perspectives at an intermediate level, between that of Case Studies 1 and 2. They spend four and a half minutes considering the fictionalized dialogue. In the first two minutes, they agree with Kristen’s gradual change idea and consider the possibility that Amara’s instantaneous idea could be relevant depending on the choice of timescale. For the remaining time, they draw the speed-time graph for a gradual change.

After reading the prompts silently, Penny initiates discussion when she asks her partners if they believe the speed increases gradually, over a short period of time, or changes in an instant. Penny shares her experience of perceiving the speed change as instantaneous, but also seems to consider the possibility that the change is gradual, if timescale is accounted for (“visually it was an instant but technically time-wise...”). Brian states, without explanation, that he thinks the speed changes gradually, and not in an instant (“it would be like a gradual just like very quickly gradual, but I don’t think it’s like an instant though”). In response, Penny elaborates on her idea of timescale, “I guess it also depends on how much within those 4 seconds how much of time are you counting, like tenths of a second, hundredths of a second...” She then turns to Cami and points to the blank graph on the laptop screen to explain what the graph would look like for a gradual change when considering smaller time intervals. This phase of the discussion ends when Huy asks, “Who did you agree with?” For the remainder of the episode, the group considers how to draw a speed-time graph for a gradual change that occurs over a four-second time interval.

We categorized Case Study 3 as moderate level EAP because the students discuss, in some detail, both the gradual and instantaneous change perspectives. The students do not explore why the instantaneous perspective is flawed, but do discuss how a gradual change could look like an instantaneous change if the time intervals on the graph are large
enough. Huy’s prompt to decide which fictitious student they agree with seems to curtail further exploration; the students quickly settle on the idea of a gradual change, and spend their remaining time working on that graph.

V. DISCUSSION AND CONCLUSION

The activity involving the FSD between Amara and Kristen elicited high-level EAP in two groups, moderate EAP in three groups, and low EAP in three groups. All groups discussed both the gradual change perspective of Kristen and the instantaneous change perspective of Amara, consistent with the activity prompts, which asked students to sketch the speed-time graphs for both perspectives. We speculate this FSD led to more EAP than an activity that presents only a single, normative perspective linked closely to the correct answer to a specific question or task. Such an activity, with a relatively less open-ended format, would seem likely to lead most students to consider only a single solution. Future research could investigate the extent to which FSDs, as a form of instruction, do in fact enhance student exploration of alternative perspectives.

This activity’s explicit directions to produce graphs, in tandem with students seeking to identify the “right” idea, may have interfered with student engagement with the two perspectives. In comparison to the moderate and low EAP groups, students in the two high EAP groups engaged in longer discussion of the two perspectives before shifting attention to the graphs. In the low and moderate EAP case studies, the conversations shifted relatively quickly from more open-ended consideration of the contrasting perspectives to more directed efforts to agree on a correct answer. For example, Huy asks “Who do you agree with?” in the moderate EAP case study, and Amy states “Now we have to sketch the speed time graph” in the low EAP case study.

All three case study groups discussed timescale. The low and moderate EAP groups did so when drawing their graphs. The moderate EAP group (Case Study 3) focused on the scale on the time axis for the graphs, while the low EAP group (Case Study 2) focused on the length of time over which the change occurred, with less focus on the scale of the time axis. Regardless of which perspective a group endorsed, groups noted a perception that the speed of the cart changes “immediately” when the finger taps it. The high EAP group’s discussion focused on the real-world example they had generated: driving a car. We speculate that this example was motivated in part because the changes in speed of a car occur on a more perceptible timescale than the changes in speed of a lab cart that is given a quick tap while moving on a track. The high EAP group focused on shorter lengths of time, 0.1 and 0.01 seconds, over which the car’s speed changed. We infer that they recognized the differences in time scale between the car and the cart contexts, and examined shorter time scales as a way to translate from their real-world experience of the car to the laboratory experience of the cart. Contrasting the familiar example of the car with the more “physics-y” example of the cart may have supported the high EAP students in abstract thinking about timescale and the nature of changes in speed. Investigation of student exploration of alternative perspectives in specific FSD activities can identify productive resources, such as the use of driving a car as a real-world example of how speed changes over time, that are activated in groups with high-level EAP. These resources could then be available to instructors as tools to increase student consideration of alternative perspectives.

The high EAP group of Case Study 1 engaged in the productive but uncommon practice [18] of disconfirming an alternative explanation. They could, however, have engaged in even further exploration of the two perspectives. For example, the group could have generated novel examples of changes that are instantaneous in nature (e.g., the change in the amount of a bank account when a paycheck is deposited, or the change in the energy of an atom when a photon is emitted). Questions that are more esoteric, but still meaningful (e.g., whether falling in love occurs gradually or suddenly) might also foster higher levels of exploration of alternative physics explanations and perspectives. Supplementing this FSD with an activity to engage students in contrasting cases of instantaneous and gradual changes beyond speed could help students to engage more with the alternative perspectives. Instructors could use such strategies across the modules in the NGP curriculum.

Many PER-based materials use FSDs to engage students with contrasting perspectives [1–3]. In the FSD studied here, most groups did engage in moderate to high levels of exploring alternative perspectives. Exploring alternative perspectives, however, is only one of many communication patterns that may arise in a social learning environment. Further study can investigate whether FSDs foster other productive learning behaviors aligned with the socio-metacognitive framework. This framework provides a structure for investigating how FSDs support student collaboration for synthesizing information and negotiating new knowledge in equitable and constructive discourse. Further work on FSDs can identify what effects they have on learning dynamics in small group environments and how to design and integrate them into curricula to maximize their benefits.

ACKNOWLEDGEMENTS

This material is based upon work supported by the NSF, DUE-2021547 and DUE-2021307. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF. We also thank Carolina Alvarado.
[7] M. Kryjevskaia, P. R. Heron, and A. F. Heckler, Intuitive or rational? students and experts need to be both, Physics Today 74, 28 (2021).
Adding self-regulated learning instruction to an introductory physics class

Danielle Maldonado (she/her/hers) and John Stewart (he/him/his)

Department of Physics and Astronomy, West Virginia University, 135 Willey St., Morgantown, West Virginia, 26506

Self-regulated learning (SRL) is an essential factor in academic success. Self-regulated learning is a process where learners set clear goals, monitor progress toward attainment of those goals, and adapt their strategies to improve their learning. Because SRL is often not explicitly integrated into the classroom, students struggle to identify and use learning techniques empirically proven to be more successful than others. SRL is a learned skill students can develop over time that has been found to be related to high achievement and self-efficacy. This paper examines the effects of introducing SRL strategies into an undergraduate introductory physics classroom. The degree to which the students were self-regulated learners was correlated with their test averages ($r = 0.23$, $p < 0.05$). Students reported that they found the SRL instruction helpful (3.5 out of 5.0 on a 5-point scale) and 86% of the students felt the time spent on the instruction was generally appropriate. Students’ preferred study methods changed over the course of the semester, indicating that students applied SRL by adapting their learning processes based on which methods were most effective in helping them study for an upcoming exam and opting not to use techniques no longer perceived as useful. Higher achieving students were more likely to settle on highly effective techniques by the end of the semester, while lower achieving students continued to modify their learning processes.
I. INTRODUCTION

Self-regulated learning (SRL) is a skill that learners develop to direct their cognition, metacognition, behaviors, and performance to achieve their academic goals [1]. While there are many theoretical models for self-regulated learning, several share a few common features. First, goal-setting begins the self-regulated learning process by outlining the objectives of the learning period [2]. This includes choosing learning strategies dependent on their effectiveness at achieving the desired learning outcomes. Next, self-regulated learning is a cyclical process involving feedback loops that inform learners on how to proceed toward their goals [3]. Learners metacognitively monitor the effectiveness of their chosen strategies, distractions from internal or external sources such as the environment, and motivation levels throughout the learning period. Finally, self-regulated learners reflect on the successes and failures of their learning processes [4]. This includes identifying successful or unsuccessful learning strategies and measuring the attainment of their previously set goals. Self-regulated learners determine the strengths and weaknesses in their learning regimens, make necessary changes toward goal achievement, and the cycle continues.

In a study on self-regulatory processes, 93% of students were correctly classified as higher achieving or lower achieving based on their knowledge of self-regulated processes; the use of SRL strategies predicted their standardized test scores after controlling for non-cognitive variables [5]. However, in another study, when asked to rank the general effectiveness of learning strategies, undergraduates were unable to identify the methods which have been empirically determined to be more effective [6]. SRL is a learned process that helps students develop effective study strategies [7] and has been previously implemented in physics courses with a positive effect [8]. Because the use of SRL strategies has been shown to increase student grade point averages [9] and be correlated with academic achievement in the natural sciences [10], it is likely that students would benefit from instructors incorporating SRL instruction into their classrooms.

This study presents initial results of the implementation of SRL instruction in undergraduate physics classrooms. The following research questions will be explored: RQ1: Does the self-regulated learning intervention produce positive student results? RQ2: How do students’ preferred study strategies and behaviors change over time when exposed to self-regulated learning instruction?

II. METHODS

This study was conducted at a R1 university with greater than 18,000 undergraduate students as of Fall 2021. The institution’s undergraduate population was 81% White, 6% two or more races, 4% Hispanic/Latina, 3% Black or African American, 3% non-resident alien, 2% Asian, and other groups less than 1% [11]. The intervention program was administered over two semesters from Fall 2022 to Spring 2023 in the introductory, calculus-based electromagnetism course taken by physical scientists and engineers. The class features three 50-minute lectures each week which use Peer Instruction [12] and group problem solving and one 170-minute lab each week which implements a number of active learning strategies. Students were presented with ten SRL and metacognitive techniques during the lecture segment of the course as well as a general discussion of SRL and metacognition and why it was important in learning physics. One new technique was presented each lecture through the first half of the class using one to three powerpoint slides with discussion occupying approximately two to three minutes of class time.

The theoretical model for the study was based on models developed by Zimmerman [13] and Pintrich [14]. The SRL program was developed around a cyclic application of “Plan,” “Monitor,” and “Reflect.” Students were exposed to new learning techniques they may not have been familiar with and encouraged to experiment with implementing these strategies into their study routines. The strategies presented included six SRL techniques: (1) Take Care of Yourself, (2) It’s Never Too Early, (3) Study A Little A Lot, (4) Don’t Rote Memorize, (5) Quiz Yourself, and (6) If At First You Don’t Succeed, Try Something Else. These strategies were chosen based on previous work by Lineweaver on the memorability and perceived positive influence of the techniques [15]. Additionally, students were presented with four metacognition techniques: (1) Concept Maps, (2) Take Notes From Memory, (3) Think Out Loud, and (4) Identify the Muddiest Point [16, 17].

To encourage the key reflection and planning parts of the SRL cycle, all students were sent four online surveys; they received a small amount of course credit for completing each survey (10 total bonus points out of 1000 total class points). Only students consenting to participate in this research study (99.1% of students taking the surveys) are included in the following analysis. The first survey was sent the week before the first exam and asked the students to select the techniques they planned to use on the upcoming exam and to rate their expected effectiveness of the study techniques. A large list of possible study methods were presented; a subset of that list is shown in Fig. 1. The last three surveys were given one week after the first three in-semester exams. These surveys asked the students to reflect on their previous exam performance and rate the study methods by effectiveness. The students reported the most and least effective techniques through open-response answers as well as which techniques they would add or remove from their study plan.

On the last survey, students were asked about the effectiveness of the SRL instruction overall to measure the students’ perceived value of the program. In Spring 2023, Survey 4 also included additional questions measuring the perceived effectiveness of specific SRL and metacognitive techniques. In Spring 2023, Survey 4 also contained 17 questions to assess the degree to which students were “self-regulated learners.” These Likert-scale questions were developed from items in the Motivated Strategies for Learning Questionnaire (MSLQ).
The items measured a number of constructs important to SRL including goal setting, environment and time management, cognitive/metacognitive strategies, help-seeking, and evaluation.

III. RESULTS

Table I presents the descriptive statistics for the two semesters: the overall class enrollment \((N)\), the number of students who completed each survey \((n)\), and the test average of the 3 in-semester examinations. The average survey response rate was \(81\%\) rate in the fall and \(76\%\) in the spring. While response rates decreased for later surveys, they remained high throughout the semester.

<table>
<thead>
<tr>
<th>Semester</th>
<th>Fall 2022</th>
<th>Spring 2023</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = Total</td>
<td>239</td>
<td>136</td>
</tr>
<tr>
<td>n = Survey 1</td>
<td>196</td>
<td>121</td>
</tr>
<tr>
<td>n = Survey 2</td>
<td>203</td>
<td>104</td>
</tr>
<tr>
<td>n = Survey 3</td>
<td>192</td>
<td>95</td>
</tr>
<tr>
<td>n = Survey 4</td>
<td>186</td>
<td>95</td>
</tr>
<tr>
<td>Test Average ((1 - 3))</td>
<td>74.6 (\pm) 17.8</td>
<td>73.2 (\pm) 19.1</td>
</tr>
<tr>
<td>DFW Rate</td>
<td>14%</td>
<td>15%</td>
</tr>
</tbody>
</table>

A. Intervention Efficacy

To provide evidence for the value of SRL instruction and its potential benefits for students, we measured the degree to which students implemented self-regulated learning behaviors late in the course using the items developed from the MSLQ [18]. The degree to which students were self-regulated learners by Survey 4 was positively correlated with their average on Exams 1 to 3 with correlation \(r = 0.23\), \(p < 0.05\). This represents a small to medium effect by Cohen’s criteria [19]. This correlation represents a stronger association between a non-cognitive variable and academic achievement than a number of other constructs thought to influence achievement, such as test anxiety, the student’s personality, or their goal commitment [20]. The correlation is weaker than that of self-efficacy or cognitive measures such as ACT scores with academic achievement.

The DFW rate – the fraction of students earning grades of D or F or withdrawing from the course – was lower in the semesters implementing SRL instruction than in the two semesters prior to introducing SRL instruction. The DFW rate in Fall 2021 was \(16.5\% \pm 7\%\), which decreased to \(14.0\%\) in Fall 2022. The DFW rate in Spring 2022 was \(22.5\% \pm 6\%\), which decreased to \(15\%\) in Spring 2023. Historically, the DFW rate for the class is smaller in the fall which is taken by more on-sequence students.

The students were asked to rate the effectiveness of the SRL and metacognitive techniques presented on a 5-point scale \((5 = \text{most effective})\) as shown in Table II. Of the 10 tips presented, reminding students to Take Care of Themselves was reported as the most effective. This may suggest that simple messages promoting mental and physical health and well-being can be productive when included in physics classes. Students also reported that the advice to start their assignments and exam study sessions earlier, as opposed to cramming, was beneficial. This result was supported by a qualitative analysis of the open-response answers; the most common strategies students wished to add to their study regime included starting earlier and improving time management (Table IV).

The students rated the helpfulness of the in-class SRL discussion as \(3.52 \pm 1.0\) on a 5-point scale; \(57\%\) of participating students felt that in-class discussion was somewhat or very helpful. The helpfulness of the reflection into past study habits provided by the surveys was rated \(3.55 \pm 1.1\); \(59.8\%\) found the reflection as somewhat or very helpful.

The time spent discussing SRL reduced the time spent discussing physics content. The most direct measure of the student’s perceived value was whether they thought the time was well used. Table III shows the results of a question asking how the time used by SRL instruction should be modified for
FIG. 1. Figure (a) Study methods and their perceived effectiveness students thought they would use to study for Exam 1 ($n = 317$); Figure (b) methods and their perceived effectiveness students did use to prepare for Exam 3 ($n = 281$).

The overwhelming majority, 86%, of the students felt the SRL instruction used about the right amount of time (Table III either keep the same, reduce slightly, or increase slightly).

B. Patterns of self-regulation

The survey items asking students to reflect on their study habits allowed the investigation of how student perceived their habits to have changed through the semester and gave some indication of what level of self-regulation could be expected.

Figure 1(a) presents student responses to the question, “Select all the study methods you plan to use for Test 1 and how effective you think they will be. If you don’t plan to use the method, select Will not do.” Students were asked about many methods, a subset of those methods are presented in the figure. These include “Review lecture notes,” “Rewrite lecture notes,” “Review homework solutions,” “Rework homework,” “Work practice tests,” “Read textbook,” “Review labs,” “Attend office hours,” and “Study with others.” Students had access to their instructor’s class lecture notes and three practice tests for each exam.

Before Exam 1, students reported that most of the study methods presented would be effective to some degree in their studying and indicated they would use those methods. About 67% of participants (213 of 317) believed that completing the practice tests would be “extremely effective,” while other methods such as reviewing homework solutions and reading the textbook were more often perceived as “very effective” or “somewhat effective.” While some methods stood out as likely to not be utilized by students, overall, most methods were viewed as likely to be effective prior to the first exam of the course.

Figure 1(b) presents student responses to the question, “Select all the study methods you used for Test 3 and how effective you think they were. If you didn’t use the method, select Did not use.” The figure shows a dramatic shift in habits and their perceived effectiveness from those in Fig. 1(a). Working through practice exams remained the most favored method by students for the third exam, but there were several more techniques than previously that students were opting not to include in their studying. 48% of students (134 of 281) would not study with others, 63% (177 of 281) would not rewrite lecture notes, and 76% (213 of 281) would not attend office hours. Students continued to utilize the course textbook throughout the semester; however, it was often viewed as a less effective tool compared to other study techniques. This agreed with students’ qualitative open responses that frequently listed using the textbook as the “least effective” study method (Table IV). The change from Fig. 1(a) to (b) happened over the semester; Surveys 2 and 3 showed a progressive change in study methods suggesting students were reflecting on their past performance and modifying study behavior accordingly.

The prior analysis aggregated all students, but one might expect students who are doing well in the class to modify their study behaviors less than students who are struggling. To analyze the differences between higher achieving and lower achieving students, participants were asked to report their most effective and least effective study method in an open-response question and asked to identify any strategies they intended to add or remove from their study plan. Higher achieving students were students with a test average above the 75th percentile after Exam 3, while lower achieving students were those with a test average lower than the 25th percentile. A qualitative analysis of these responses is presented in Table IV. Both groups of students reported that reviewing the instructor-provided practice exams and home-
TABLE IV. Study strategies of higher and lower achieving participants after Exam 3.

<table>
<thead>
<tr>
<th>Most Effective</th>
<th>Least Effective</th>
<th>Add Strategy</th>
<th>Remove Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Strategies of Higher Achieving Participants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(75th Percentile) after Exam 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51 - Practice exams</td>
<td>23 - None</td>
<td>69 - None</td>
<td>70 - None</td>
</tr>
<tr>
<td>23 - Reviewing HWs</td>
<td>15 - Readings</td>
<td>1 - Focus on free-response</td>
<td></td>
</tr>
<tr>
<td>11 - Readings</td>
<td>12 - Reviewing notes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 - Reviewing labs</td>
<td>(+20 responses / 10 strategies)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 - Reviewing notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(+15 responses / 9 strategies)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study Strategies of Lower Achieving Participants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(25th Percentile) after Exam 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42 - Practice exams</td>
<td>22 - Readings</td>
<td>55 - None</td>
<td>3 - Procrastination</td>
</tr>
<tr>
<td>11 - Reviewing HWs</td>
<td>15 - Not studying enough</td>
<td>5 - More practice problems</td>
<td>2 - Remove distractions</td>
</tr>
<tr>
<td>8 - Reviewing labs</td>
<td>7 - Reviewing notes</td>
<td>4 - More time studying</td>
<td>2 - Readings</td>
</tr>
<tr>
<td>5 - Readings</td>
<td>7 - None</td>
<td>2 - YouTube videos</td>
<td>1 - Reviewing notes</td>
</tr>
<tr>
<td>5 - Studying w/ others</td>
<td>6 - Lectures, lecture slides</td>
<td>(+ 12 responses / 10 strategies)</td>
<td></td>
</tr>
<tr>
<td>5 - None</td>
<td>(+9 responses / 7 strategies)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(+16 responses / 6 strategies)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Work were the most effective study methods. Both groups also agreed that reading the textbook was the least effective. 23 of 70 high-achieving students reported that none of their study methods were least effective. No higher achieving student said they would remove any methods from their studying, and only one noted they would add to their study plan. In contrast, lower achieving students listed a variety of methods they would add or remove from their processes, including spending more time on studying and removing distractions and cramming.

**IV. DISCUSSION AND CONCLUSIONS**

**RQ1:** Does the self-regulated learning intervention produce positive student results? Students found the SRL instruction was helpful and felt the 2-3 minutes per lecture spent on the instruction was appropriate. “The study habits helped me because I knew physics was a hard subject for me and I didn’t really understand how to study for it,” one student commented. Higher levels of self-regulating behaviors were correlated with higher exam scores ($r = 0.23$, a small to medium effect), which agrees with past work on SRL and achievement [5, 9, 10]. The DFW rate in the semesters implementing SRL instruction was lower – substantially lower in the spring semester – than in the prior year’s offering of the class. The participation rate for students was high from 75% to 80% with small incentivization in the form of course credit. As such, SRL instruction is a promising, low-cost method to help students succeed in introductory physics.

**RQ2:** How do students’ preferred study strategies and behaviors change over time when exposed to self-regulated learning instruction? Throughout the semester, students self-regulated by actively modifying their study habits and choice of techniques between each exam. Figure 1 shows a dramatic shift from students’ planned behavior and its expected effectiveness before Exam 1 to their actual behavior and its perceived effectiveness when reflecting on Exam 3. This suggests students are practicing SRL by planning to employ a study method, monitoring its usefulness, and reflecting on its effectiveness. Beyond methods reported in Fig. 1, students were also asked to rate the effectiveness of reading online physics material, watching online physics videos, and using tutoring sites. Before Exam 1, approximately 61% of students reported using these online sources to some degree. By the end of the semester, only about 25% of students reported using these resources, suggesting students recognized that using the materials provided by the class, particularly the practice tests, were more useful. The patterns of self-regulation differed for higher achieving and lower achieving students. Higher achieving students established a study regime using methods they found effective by the end of a semester, while lower achieving students struggled to develop stable study plans and continued to make modifications late in the semester.

This study has potential limitations. No control group was available for the study. The data were collected at one institution with predominantly white students. Demographic variables such as gender and race were not collected; therefore, it is unknown if the program served different students to different degrees.

A program of SRL instruction supported by online surveys to promote reflection and planning represents a low instructional time-cost method to help students develop self-regulation skills. This work is supported in part by National Science Foundation grant DUE-1833694.
Upper-level students’ conceptual understanding of energy and momentum

Alexandru Maries (he/him/his)
Department of Physics, University of Cincinnati, Cincinnati, OH, 45220

Mary Jane Brundage (she/her/hers), and Chandralekha Singh (she/her/hers)
1Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA, 15260

The Energy and Momentum Conceptual Survey (EMCS) is a multiple-choice survey that contains a variety of energy and momentum concepts at the level of introductory physics used to help inform instructors of student mastery of those concepts. Prior studies suggest that many concepts on the survey are challenging for introductory physics students and the average student scores after traditional instruction are low. The research presented here uses the EMCS to investigate the extent to which upper-level students have developed mastery of energy and momentum concepts both before and after an upper-level classical mechanics course. To contextualize their performance, it will be presented alongside the performance of introductory students. Additionally, a different set of upper-level students provided explanations for their answers to each question, and those explanations were useful for understanding their common difficulties with energy and momentum concepts. We discuss some of the most challenging questions on the EMCS for upper-level students and common reasons they had difficulty with those questions.
I. INTRODUCTION AND GOAL

One goal of an introductory physics course for physical science and engineering majors is for students to develop a robust conceptual understanding of the underlying physics [1–15]. This remains one of the goals in courses for physics majors as they take advanced physics courses even though their courses use complex mathematics to solve problems [16]. In these courses, students are often expected to make the math-physics connection themselves to develop a good knowledge structure without much scaffolding support from instructors [16, 17]. In particular, the focus of upper-level undergraduate courses is often on being able to solve complex quantitative problems with course assessment focusing exclusively on such problems, which students can do without a robust understanding of the underlying concepts.

In the context of energy and momentum, students often learn about the basic concepts including the work-energy theorem, conservation of mechanical energy, impulse-momentum theorem and momentum conservation in the calculus-based introductory courses. Moreover, in the typical traditionally taught upper-level undergraduate mechanics courses, students are taught these same concepts to solve complex quantitative problems. However, these problems can often be solved algorithmically without any course assessment specifically focusing on conceptual understanding. If advanced students in these traditionally taught courses are not given grade incentives and provided scaffolding support to make the appropriate math-physics connections, they may solve problems by pattern matching without necessarily unpacking the underlying concepts for developing a deep conceptual understanding of the foundational concepts, which is important for cultivating expertise in physics [18-20].

It is therefore useful for instructors to be aware of the extent to which upper-level students have mastered foundational concepts of energy and momentum, and the extent to which students improve their conceptual understanding after a traditionally taught course in classical mechanics primarily focused on helping students develop mathematical prowess. A few conceptual surveys have been developed for energy and momentum concepts, namely, the Energy Concept Assessment [21] and the Energy and Momentum Conceptual Survey, or EMCS [7]. In this study, we use the EMCS to study the persistent difficulties of upper-level undergraduates students with concepts measured by the EMCS as follows:

RQ1: On which EMCS questions do upper-level students struggle after instruction? Are there any patterns when comparing student responses in introductory and upper-level courses for these questions?

RQ2: What are the common difficulties of advanced students based upon the alternative choices they commonly selected? Are they the same or different from the alternative choices selected commonly by the introductory students?

II. METHODOLOGY

As mentioned, we use the EMCS to investigate the performance of upper-level students. To contextualize this performance, it is reported along with the performance of introductory students at the same institution. The upper-level students were taking a required upper-level physics classical mechanics course and completed the EMCS before (first few weeks) and after traditional instruction (last few weeks) while the introductory students did the same except at the beginning and end of their introductory physics course. Two years of data were collected for the upper-level students. We note that in the data shown in Table 1, some students took the pre-test but not the post-test and vice versa. We have carried out the analysis using matched data only (all students took both the pre-test and the post-test) and found very similar results. Therefore, we elected to report the analysis for the unmatched data here.

Additionally, 26 upper-level students from a different class completed the EMCS and provided explanations for each answer. We used these explanations to identify common reasons that students struggle on some of the questions on the EMCS and we discuss in detail the most challenging six questions.

III. RESULTS

Table 1 shows the performance of introductory students and upper-level students for each question on the EMCS. Since our focus is on upper-level students, their improvement as measured by normalized gain, g, [22] defined as \((\text{post }\% - \text{pre }\%)/(100 - \text{pre }\%)\), is listed and the questions on which they performed well (performance above 80%) are highlighted in green and those in which they performed poorly (performance lower than two thirds) are highlighted in red.

To answer RQ1, we identified questions on the EMCS on which the performance of upper-level students was below two thirds and also compared their performance with that of introductory students to identify any patterns. We found that on roughly half the questions on the EMCS, the performance of upper level students is below two thirds, namely Q3, Q5, Q6, Q8, Q9, Q10, Q12, Q13, Q16, Q17, Q22, Q23, Q24. This suggests that the upper-level students are far from having developed a robust conceptual understanding of the concepts of energy and momentum after traditional instruction in an upper-level classical mechanics course. Comparing the post-test performance of the upper-level students with the corresponding post-test performance of introductory students on these questions, the data show that in nearly all of the questions (11 out of 13), introductory students’ post-test performance after instruction was less than 50%. In fact, on all of the questions on which the performance of introductory students was below 50% on the post-test, the performance of the upper-level students after instruction was below two-thirds, indicating that the concepts covered in those questions are very challenging for even the upper-level students to grasp after traditional
instruction, and instructors need to do more to help students develop a solid understanding of the underlying concepts using evidence-based approaches.

Furthermore, the data in Table 1 shows that on all the questions on the EMCS on which the post-test performance of upper-level students was high (better than 80%), their pre-test performance was greater than 70%. Also, normalized gains tell a similar story: all the highest normalized gains are on the same questions, and part of the reason for it is the high pre-test performance. For example, Q11 which showed the highest normalized gain of 0.68 was due to an increase from 85% to 95%. More importantly though, on many of the questions with low post-test performance, student performance exhibited very low or negative normalized gain. For example, Q8, with an improvement from 57% to 62%; a normalized gain of 0.11. These data point to the same interpretation: whenever student pre-test performance was below a certain level, it is very difficult for students to improve much even in an upper-level classical mechanics course.

To answer RQ2, we identified the most common incorrect answer choices in each group by looking at the percentage of students who select each answer. Then, we identified questions in which the most common incorrect answer choices differed between introductory and upper-level students. We found that in most cases, the most common incorrect answer choices of upper-level students were the same as for the introductory students, except the percentages were lower in the upper-level student group. To illuminate some difficulties students had on the EMCS, we focus on the six questions that were most challenging for upper-level students. We note that all the questions can be found in the appendix of Ref. [7] and that for each concept, the EMCS probes student understanding in multiple contexts. Generally, student performance is context dependent, which means that they are likely to struggle in certain contexts but not in others because they are yet to have developed expertise. Our goal here is to focus on the specific contexts in which even upper-level students struggled.

Q3 describes a situation in which a white hockey puck collides elastically with a red hockey puck (and no net external forces act on the two-puck system) and asks which statements are correct among (1) the kinetic energy of the white puck is conserved, (2) the linear momentum of the white puck is conserved, and (3) the linear momentum of the two-puck system is conserved. The only correct statement is statement (3); however, 36% of upper-level students selected statements (1) and (3), due to only focusing on the fact that kinetic energy is conserved and not recognizing that this is only true for the two-puck system in this context. For example, one student stated “No external forces so momentum of the system is conserved, and it is elastic so kinetic energy is conserved.” What is interesting is that these students can clearly recognize that it’s the linear momentum of the two-puck system that is conserved, not just the white puck, and this doesn’t appear to be sufficient to help them recognize that the same must be true for kinetic energy conservation. It is possible that students have formed a strong association between “elastic collision” and “kinetic energy conservation” and in this context, have neglected to consider the applicability conditions (i.e., that it is the kinetic energy of the system is conserved), which is a common student difficulty in physics [7,19,20].

Q3 is also intriguing because it’s the only question on the EMCS with a significant decrease in performance of 17% from 65% on the pre-test to 48% on the post-test. This is the only question which had a decrease larger than 10%. Additionally, it is the only question in which the upper-level students had a lower performance than the introductory students (48% compared to 61%), and also, more upper-level students thought that the kinetic energy of the white puck is conserved (36%) compared to introductory students (21%). This suggests that the low performance of upper-level students might be due to them not paying close attention to the fact that in a collision, it is the kinetic energy of the system that is conserved and it points to the importance of emphasizing to students whenever conservation laws are discussed that laws such as conservation of mechanical energy and conservation of momentum are always applied to a system of multiple objects and it is the total energy/momentum of the system that is conserved, not that of an individual object.

Q5 relates to the impulse momentum theorem: two bullets are shot towards blocks of equal mass, one made of wood and the other made of steel. The bullet embeds in the wood block and bounces off of the steel (elastically), and students are asked which block travels faster after the collision. Only roughly half of the upper-level students answered this question correctly.
correctly and 29% of students selected that the wood block would travel faster because the bullet transfers all of its kinetic energy to it. For example, one student stated that “the wood block now has the KE of the bullet instead of splitting it between two objects,” and another stated that “all of the bullet’s energy goes into the wood, while the bullet is still moving for the steel, that implies the velocity of the wood block will be more.” These students are not recognizing that when the bullet embeds in the block, some of its kinetic energy is used to deform the block and lost to heat, so the kinetic energy of the system is not constant. In other words, students expect that both the kinetic energy and the momentum of a system is conserved in a perfectly inelastic collision. We also note how strong this alternate conception appears to be in this context: the percentage of students who have this alternate conception is the same in introductory and upper-level courses (30% for introductory students compared to 29% for upper-level students).

Additionally, analyzing students’ choices more carefully, introductory students selected that the wooden block would be moving faster after the collision for two reasons: 1) because it gains the momentum of the bullet (20%), and 2) because it gains the kinetic energy of the bullet (30%). Upper-level students who expected that the wooden block would be moving faster primarily selected the first reason, and the percentage of upper-level students who selected it (29%) was comparable to that of introductory students (30%).

Q23 is similar to Q5 in that it also relates to the impulse-momentum theorem: two balls, one made of rubber and the other made of putty, are dropped from the same height. The rubber ball bounces up after the collision with the ground while the putty ball comes to rest. The question also states that the collisions take the same amount of time $\Delta t$ and asks which ball exerts a larger force on the floor. On this question, 29% of upper-level students stated that the forces are the same. For example, one upper-level student explained: “Average force applied by the surface will equal the change in momentum / $dt$. It is the same in both cases.” Also, students seem to answer Q5 and Q23 quite consistently: roughly equal percentages of students answer the questions correctly (52% and 55% on Q5 and Q23, respectively), and the percentages of students with the aforementioned alternate conception is nearly identical as well (29% on both questions).

Q6, the second most challenging question on the EMCS, provides a diagram showing the circular orbit of a satellite around the Earth (Fig. 1) and asks which statement is correct among (a) the gravitational potential energy of the satellite decreases as it moves from A to B, (b) the work done on the satellite by the gravitational force is negative for the motion from A to B, (c) the work done on the satellite by the gravitational force is zero for the motion from A to B, (d) the velocity of the satellite remains unchanged as it moves from A to B, and (e) none of the above. Q6 is a very difficult question for upper-level students: only 38% answered it correctly, a performance not a lot higher than that of introductory students (28%). On this question, the most common difficulties of both upper-level and introductory students were the same: 31% of upper-level and 37% of introductory students selected (d) and 21% of upper-level and 24% of introductory students selected (e). Students who selected (d) did not distinguish between velocity and speed, e.g., one student who selected this answer choice motivated their choice by saying “It must move with constant velocity to have a circular path.”

In Q22 three balls are launched from the same height at the same speed but at different angles (ball 1) is launched vertically, ball 2 is launched at an angle of 60° with respect to the horizontal and ball 3 is launched at an angle of 45° with respect to the horizontal). They all reach a dotted line at a certain height above their starting position and students are asked to rank the balls according to their speed when they reach the line. Q22 is the most difficult question on the EMCS: only 29% of upper-level students answered it correctly, which is identical to the percentage of introductory students who answered it correctly. On this question, roughly one third of students rank the speeds (from largest to smallest) in the same order as their $y$ components and also one third rank them in the same order as their $x$ components. For example, one upper-level student stated: “The parameter that matters here is the vertical velocity. for [ball 1] it is the biggest, at $v_y$, for [ball 2] it is the second biggest at $v_0 \sin 60^\circ = v_0 \sqrt{3}/2$. for [ball 3] it is the smallest at $v_0 \sqrt{2}/2$, therefore the order is 1), 2), 3).” Also, sometimes, even if a student recognized that conservation of mechanical energy could be used in this context, they still appeared to get distracted by the fact that the velocities in the $y$ direction are not the same. For example, another upper-level student who selected the same order stated: “I imagine the equation $KE_1=PE+KE_2$ at the dashed line. If this is correct, then all of the masses cancel out. The fastest ball will be whichever one has the most velocity in the vertical direction.” Other students appear to have focused on the $x$ component of velocity not being affected by the force of gravity in this question, reasoning that the object with the most initial component of velocity in the $x$ direction will be the one that reaches the dashed line moving fastest. For example, one upper-level student stated: “45 degrees has the highest component of horizontal velocity which gravity cannot affect and so moves more quickly when reaching the horizontal line”, and another upper-level student stated: “Vertical speeds are killed by gravity, while horizontal speeds are not.” While it is true that ball (1), which is thrown vertically, has more vertical velocity than ball (2), which is thrown at an angle of 60°, we need to consider both components of velocity because the question is asking for the total speed. It is not obvious if one does not consider conservation of mechanical energy that the horizontal velocity of ball (2) “compensates” for the lower vertical velocity such that the total speed is the same (both initially and when reaching the dotted line), so it is not surprising that not invoking conservation of mechanical energy leads to difficulties answering the question correctly. This suggests the importance of helping students consider multiple solution approaches to identify which approach is simpler, and
also encouraging them to contemplate whether the use of conservation of mechanical energy is applicable whenever possible because it generally leads to more efficient solution, e.g., than using two-dimensional kinematics.

Q10 describes an explosion in which a bomb at rest on a horizontal surface breaks up into three fragments, all of which fly off horizontally and asks which statements are true among (1) the total kinetic energy of the bomb fragments is the same as that of the bomb before the explosion, (2) the total momentum of the bomb fragments is the same as that of the bomb before explosion, and (3) the total momentum of the bomb fragments is zero. On this question, 29% of upper-level students thought that all three statements are correct, possibly because they associate explosions with both momentum and kinetic energy conservation. For example, one student motivated choosing all three by stating “by momentum and energy conservation.” It is interesting that students who selected all three recognized that the initial momentum is zero (because the initial velocity is zero), and thus would presumably also be able to recognize that the initial kinetic energy must be zero as well in this particular context. But the final kinetic energy cannot be zero because all three pieces are moving. In other words, it is possible that some students have not internalized that kinetic energy is a scalar quantity. For example, one upper-level student who provided explanation reasoned that the kinetic energy stays the same by stating “the amount of total kinetic energy does not change, but just the orientations do. So even if it is represented different the overall [kinetic energy] will be the same.” This type of response suggests that instructors should do more at both the introductory and advanced level to help students recognize the differences between momentum and kinetic energy in the context of collisions (and explosions).

IV. SUMMARY

We find that many upper-level students struggled with introductory mechanics concepts on the EMCS after traditional instruction in upper-level classical mechanics. The questions on the EMCS are conceptual in nature and upper-level students have difficulties with several conceptual aspects of energy and momentum covered in the EMCS. For example, on the two questions on the EMCS directly related to the impulse momentum theorem, the upper-level students exhibited very low performance of roughly 50%, indicating that they either need more support to recognize the utility of the impulse momentum theorem to understand that situation or they need more support on correctly interpreting the impulse momentum theorem. Additionally, we find that there were no questions on which the performance of upper-level students was higher than 80% and they showed marked improvement from the pre-test to the post-test: on all the questions on the post-test which showed a final performance of upper-level students greater than 80%, their pre-test performance was greater than 70%. While it is true that the upper-level students did improve on some questions on which they started very low in pre-test, it was unlikely for them to improve beyond a certain level. Furthermore, on roughly half the questions on the EMCS, the performance of upper-level students either decreased slightly or showed little improvement (less than 5% from the pre-test to post-test). For a few of those questions, this could be explained due to very high pre-test performance, but in most, this was not the case. We also note that on the entire EMCS, the normalized gain was a mere 14%, corresponding to an increase from a 63.2% average on the pre-test to a 68.3% average on the post-test.

We find that on all the EMCS questions on which introductory students’ performance was less than 50% in the post-test (after instruction), less than two-thirds of the upper-level students provided the correct response in the post-test. It appears that on questions in which their incoming knowledge is very low, it is very difficult to improve significantly after traditional instruction in upper-level course on these concepts. We note that traditional upper-level classical mechanics teaching and assessment did not focus on the kinds of questions that are on the EMCS, although the course deals with the same concepts. Additionally, we find that the upper-level students often displayed the same types of difficulties on the EMCS questions as introductory students, usually with lower percentages, but there wasn’t a significant difference between the two groups on some of the questions. These findings suggest that a traditional upper-level course (involving primarily lecture-based instruction and focus on quantitative problem solving) is not effective at helping students develop a robust knowledge structure of energy and momentum.

We note that we previously conducted interviews with physics instructors who had taught traditional upper-level undergraduate and graduate core courses. We found that some instructors incorrectly believe that learning physics concepts is easier for students at all levels than learning how to solve physics problems using “rigorous” mathematics [17]. They believe that if non-science majors can learn physics concepts, science and engineering majors and particularly physics majors can learn physics concepts on their own even if there was no conceptual assessment in the course and there is not much use in instructors wasting precious instructional time explicitly on concepts in advanced courses. Instructors with these types of beliefs often noted that they focus mainly on quantitative problem solving that will help students do complex calculations. The study discussed here suggests that this type of instruction is not effective at helping students develop a functional understanding of fundamental physics concepts. However, similar to introductory physics, instruction is likely to be more effective if upper-level classical mechanics courses use evidence-based approaches that integrate conceptual and quantitative aspects of mechanics. As Mazur noted [23], students can become very adept at regurgitating solution patterns using memorized algorithms, but not be able to answer ‘simpler’ questions, e.g., comparing the brightness of different light bulbs in significantly simpler circuits. If instruction does not explicitly integrate both conceptual and quantitative problem solving in teaching and assessment, students can rely on algorithmic problem-solving approaches and perform well despite lacking deep understanding of the underlying physics concepts.
Augmented Reality to Scaffold 2D Representations of 3D Models in Magnetism

Michele W. McColgan,1 George E. Hassel,1 Natalie C. Stagnitti,1 Jason W Morphew,2 and Rebecca Lindell3

1Department of Physics & Astronomy, Siena College, 515 Loudon Road, Loudonville, NY, 12211
2School of Engineering Education, Purdue University, 701 West Stadium Avenue, West Lafayette, IN 47907
3Tiliadal STEM Education: Solutions for Higher Education, Lafayette IN 47907

We report on the initial implementation of the MARVLS augmented reality app in three sections of introductory physics courses at Siena College. The courses adopted models with emphasis on the relationship between 2-D and 3-D representations of topics in magnetism. This app is intended to provide engaging and interactive visualizations of abstract concepts. The AR approach has a unique capability to illustrate relationships between physical and mathematical representations. The models were displayed on foam Merge cubes, allowing students to manipulate and explore the 3D representations. The app offers interactive features to control model components, animations, and connections to 2D representations and equations. We compare pre- and post-assessment data and note modest gains for electromagnetism concepts, and in spatial reasoning for the calculus-based physics sections. Further analysis indicates moderate improvement in specific magnetism questions, and a general indication by Likert scale that students found the AR materials helpful to understand magnetism.
I. INTRODUCTION AND MOTIVATION

We describe a pilot study on the use of a smartphone app to enrich lessons in magnetism, first formally adopted in the spring of 2023 in three introductory physics sections at Siena College. The overall goal of this on-going study is to implement activities that focus on four topics that strongly involve 3D visualization: magnetic field, magnetic force, Ampere’s Law, and electromagnetic induction. Implementation involves ensuring ease of student use, development and revision of activities, developing the framework for analysis of student engagement, and the identification of next steps for improvement. This study operates within the guidelines of a human subjects study that requires student anonymity. The classroom activities involve drawing to compare 2D and 3D representations, and answering scaffolded and open-ended questions to investigate models. Analysis includes review of the drawings and activity responses, observing student participation and collecting feedback, assessments, and conducting interviews. The major question is to determine if the models improve student comprehension of abstract concepts that require 3D visualization.

Motivation: 2D to 3D Representations Learning within STEM entails developing robust conceptual understandings and theoretical models involving abstract and complex concepts. Cognitive science is converging on the fact that our understanding of abstract concepts is grounded in embodied experiences and representations [1]. Two critical visuospatial skills critical for learning in STEM are the ability to visualize and manipulate interactions in three dimensions, and the ability to represent and understand these representations in multiple modalities (i.e., representational fluency) [2]. Many cognitive psychologists espouse that poor visuospatial skills represent the true barrier for success in STEM fields [3] (and references therein). This barrier may be higher for female students, as spatial reasoning ability is one of the only areas of cognition where a gender difference persists [4]. Spatial reasoning skills can be improved through training courses that develop general spatial visualization skills through different activities including drawing 3D representations [3]. While beneficial, this increases the course load for students with less developed spatial reasoning skills. AR represents a technology that can facilitate spatial reasoning skill development with much lower costs than other technologies such as VR. Recent research has demonstrated that the use of an AR App in addition to traditional instruction in spatial reasoning, led to additional improvements in spatial reasoning, and helped narrow the gender gap in spatial reasoning [5]. However, this still requires students with low spatial reasoning to undertake additional training outside of their typical coursework. An underexplored area of research is whether content specific AR visualizations like those in MARVLS can transfer to gains in domain general spatial reasoning tasks. If so, this would represent a pathway to develop scientists and engineers from underrepresented backgrounds while not requiring additional load on these students.

II. MARVLS: MANIPULABLE AUGMENTED REALITY VISUALIZATIONS TO LEARN SPATIALLY

MARVLS is a smartphone app developed by the first author that uses augmented reality to allow learners to interact with three-dimensional (3D) visualization of concepts in physics courses. The app helps students to develop content knowledge, representational fluency, spatial visualization, and provides opportunities for the co-creation of knowledge within the same activities saving time and making classroom experiences meaningful and impactful. The MARVLS app is available for free on the App Store for iPhones and iPads and on the Google Play Store for Android phones and tablets. QR codes for the app are provided in Figure 1. A link is provided in the app to download and print a paper cube template that can be cut out and assembled. The app includes approximately 80 augmented reality models on topics in electrostatics, circuits, magnetism, induction, optics, mathematics, and several applied physics topics. For more information about the development of the app, refer to presentations and posters given at AAPT and PERC conferences in recent years [6–15].

FIG. 1. QR scannable codes to download the MARVLS app for iOS and Android devices.

III. CLASSROOM USE CASE

The pilot study has so far been conducted in two sections of the calculus-based physics course (n=58) and one section of the life sciences physics course (n=23). We chose to include all available students for this initial stage rather than further splitting this small sample to create a comparison group. The AR content supplements but does not fully replace lecture, problem-solving, reading quizzes and homework. The in-class activities include a paper handout with instructions to navigate to the augmented reality model in the MARVLS app corresponding to each activity. Students are given foam Merge cubes to be used in class [16]. When the student views the model through their smartphone or tablet, they are viewing a digital overlay of the magnetism concept onto one of the sides of the Merge cube. The students hold the Merge cube in their hand allowing the student to rotate the 3D model, change the perspective of the model, and move the model closer or
farther away to make the model larger or smaller. In each MARVLS activity, the student is asked to view the model, draw what they see, and answer some questions to further encourage the students to interact with the 3D AR models and bring important aspects of each model to the student’s attention. Presently, students work in small teams to answer the activity questions, but they use individual cubes and we collect individual drawings and responses.

The app includes buttons and checkboxes to turn on and off different components in the model, begin or pause an animation, and highlight different components in the model to connect a 2D representation to a component of the 3D model or to link a variable in an equation to a component in the model. A new development in the app is the grey box in the right bottom corner. The purpose is to provide additional ways for the students to make connections between 2D representations and the variables in equations and the 3D models they represent.

An example of one augmented reality model that links a 2D representation of the magnetic field of a current-carrying wire to its 3D AR model is shown in Figure 2. In this model, students can turn on the current (electron model or conventional current model), turn on the magnetic field, and turn on the 2D representation. They can also touch the 2D representation in the grey box on the bottom right. When touched, the arrows in the 3D model turn green to indicate how they correspond to the traditional "dot’s and x’s" representation indicating vectors out of and into the page.

A second example of a MARVLS activity is shown in Figure 3. Two 2D representations of magnetic force on a moving positive charge in a magnetic field are shown on the left. The 3D AR model of the moving charge, magnetic field, and resulting force are shown. The 2D images are added into the model and placed to match with their orthogonal views to match the 3D model.

**Student Work** To provide some examples of student work, drawings by 4 students are shown in Figure 4. For this activity, students were working with the AR model of the magnetic field of a current-carrying wire, following the guidelines described in the previous section.

One thing to notice about the student drawings is that each student showed the magnetic field circling the wire. Three of the four students included an indication of the direction of moving charges to represent current. Each of the students drew arrows close to the wire and farther away, so did not notice that the arrows have a smaller length farther from the wire. One student included the 2D representation with dots and x’s to represent into and out of the page with his 3D drawing. The student on the bottom left included arrows to represent into and out of the page in his drawing. In a drawing not included here, a student only drew 2D representations - one looking down on the end of the wire and another similar to the picture of the 2D representation in the grey box on the bottom right. However, the student also added 3D arrows to this 2D picture on the left and right to link the dots and x’s representing into and out of the page to the arrows in the 3D model. Reviewing student work, these drawings are on the right track, but they are incomplete. As we update the activities, we will incorporate more specific directions to direct student’s attention to important details in the AR models. We plan to assess all student drawings, but include these here to indicate the level of progress so far achieved.
IV. ASSESSMENT RESULTS AND DISCUSSION

The initial study has provided much data yet to be analyzed. We are in the process of conducting interviews and reviewing worksheet responses to gain further insights. Here, we examine some assessment results.

In addition to the activities, students complete the Electricity and Magnetism Conceptual Assessment (EMCA) which was developed at Siena College [17, 18]. Students completed The Purdue Visualization of Rotations Test (PVRT) [19]. For the EMCA and PVRT, students completed the assessment at the start and end of the course. The Colorado Learning Attitudes about Science Survey (CLASS) was given once after students completed all of the AR activities [20]. For each question on the EMCA assessment, students were asked to rate through a Likert scale how much using the MARVLS app with the augmented reality models influenced their answer choice. For our human subjects study, the students created a self-identifying ID which they included on the assessments and activities to link the datasets.

Summary of Quantitative Analysis For the quantitative analysis of the impact of the augmented reality models on student learning, we report on the results of the EMCA and PVRT assessments and the CLASS survey. We compared pre and post-test averages for all students, and separately for the calculus-based physics students and the life science physics students. We report normalized gains for these data sets along with standard deviations for each. We also selected the 10 magnetism questions in the EMCA assessment and calculated the average pre-test, post-test, and normalized gains for these 10 questions. Item difficulty scores were calculated for these magnetism questions and item difficulty for these same questions is provided from a previous study by the authors for comparison [17]. The Likert scale results of how the students rated the influence of the AR activities on their answer choice is presented for the 10 magnetism questions. In tables I and II, Calc and Life represents students in the calculus-based and life sciences physics courses respectively.

EMCA Assessment results for the EMCA are given in Table I. The data in this table include the average pretest score, post-test score, and the normalized gain. The equation for the normalized gain is defined as the ratio of the number of points gained between the pre- and post-test divided by the maximum possible gain [21]:

\[
\text{normalized gain} = \frac{\text{post-test} - \text{pre-test}}{N_{\text{questions}} - \text{pre-test}}
\]

where \( N_{\text{questions}} \) is the total number of questions in the assessment. The formula is calculated for each student and the average is reported in the table. Included in Table I are the results of the CLASS survey and the standard deviation. The CLASS results show that student attitudes about science are high as the average is 3 out of a maximum of 5 and above 50%. The EMCA pretest scores for our study are lower than we reported in 2015 and 2016, reflecting a changing demographic at Siena College. The post-test scores were also lower, especially for the calculus-based courses.

<table>
<thead>
<tr>
<th>Course</th>
<th>Pre STD</th>
<th>Post STD</th>
<th>Gain STD</th>
<th>CLASS STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calc</td>
<td>9.0</td>
<td>3.0</td>
<td>14.5</td>
<td>4.8</td>
</tr>
<tr>
<td>Life</td>
<td>9.0</td>
<td>2.6</td>
<td>11.8</td>
<td>3.8</td>
</tr>
<tr>
<td>All</td>
<td>9.0</td>
<td>2.8</td>
<td>13.4</td>
<td>4.6</td>
</tr>
</tbody>
</table>

TABLE I. Average Scores, Normalized Gain Values, and Standard Deviations for EMCA Pre-test, and Post-test for all questions. The EMCA Assessment includes a total of 30 questions and the CLASS survey includes 42 questions.

<table>
<thead>
<tr>
<th>Course</th>
<th>Pre STD</th>
<th>Post STD</th>
<th>Gain STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calc</td>
<td>22.2</td>
<td>4.5</td>
<td>30</td>
</tr>
<tr>
<td>Life</td>
<td>18.2</td>
<td>6.2</td>
<td>20</td>
</tr>
<tr>
<td>All</td>
<td>20.2</td>
<td>5.4</td>
<td>50</td>
</tr>
</tbody>
</table>

TABLE II. Average Scores, Normalized Gain Values, and Standard Deviations for Purdue PVRT Pre-test, and Post-test for all questions. The PVRT Assessment includes a total of 30 questions.

PVRT Assessment results for the PVRT are given in Table II. The data in this table include the average pre-test score, post-test score, and the normalized gain. The post-test scores increased with positive gains for students in the calculus-based physics course. The scores decreased for the life science physics students.

Magnetism EMCA questions This analysis focuses on the pre-test and post-test average scores for the magnetism questions in the EMCA assessment. These include questions 21 - 30 [22]. The average scores are included in Table III. For this analysis, the normalized gain was not calculated, however the gain was positive as shown by the increase in post-test score on most questions.

Average number of correct answers for questions 21-30 for the pre-test and post-test are in columns 2 and 3, respectively. Item difficulty for the dataset reported in this paper is in the 4th column and labeled as Difficulty AR and the item difficulty from an earlier source is in the 5th column and labeled Difficulty w/o(without) AR [17]. Average number of students out of 55 that selected Agree or Strongly Agree on the Augmented Reality Likert Score for each problem is in the 6th column. To further explore the impact of the MARVLS

<table>
<thead>
<tr>
<th>Course Type</th>
<th>n</th>
<th>Pre-test STD</th>
<th>n</th>
<th>Post-test STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siena Calc-based</td>
<td>33</td>
<td>2.0</td>
<td>30</td>
<td>1.7</td>
</tr>
<tr>
<td>Siena Algebra-based</td>
<td>22</td>
<td>2.3</td>
<td>22</td>
<td>1.3</td>
</tr>
<tr>
<td>Siena All</td>
<td>55</td>
<td>2.1</td>
<td>52</td>
<td>1.5</td>
</tr>
</tbody>
</table>

TABLE III. AR Magnetism Question Activities: Average Scores and Standard Deviation for pre-test, and post-test for magnetism questions 21 - 30.
activities on student understanding of magnetism topics, we calculated the item difficulty for each of the 10 magnetism questions on the EMCA. The results are show in Table IV. Item difficulty measures the fraction of students that answer the question correctly. The item difficulty is calculated as the number of correct answers divided by the total number of students who took the test:

\[ P = \frac{N_C}{N} \]

where \( P \) is the item difficulty index, \( N_C \) is the number of students that selected the correct answer, and \( N \) is the total number of students. Note that a higher value for item difficulty indicates an easier question, while a lower value reflects a harder question. The ideal value for the item difficulty index is 50% for a new assessment [23], but learning should increase the value. The average difficulty index for questions 21-30 is 0.396 with a standard deviation of 0.176. For comparison, the average difficulty index for questions 1-20 is 0.470 with a standard deviation of 0.136. This indicates questions 21-30 on magnetism are more difficult than the first twenty questions covering the concepts of electrostatics and circuits. It is interesting to note that the difficulty decreases (higher values) for questions 26, 29, and 30 suggesting that the MARVLS for those topics likely influenced student understanding of the topics of the magnetic field of a wire and the force on a charge or wire in a magnetic field.

**AR Likert Scale** The AR Likert values represent the average number of students that took the EMCA post-test and selected agree or strongly agree indicating that the AR activity influenced their answer choice. 52 students took the post-test. Any AR Likert score over 26 indicates that more than half of the students taking the post-test selected agree or strongly agree. The Likert score for problem 21 is shown in Figure 5. It is obvious from the bar chart that more students agreed with the statement that the AR model activity influenced their answer choice for question 21 than students who disagreed or were neutral. This is indicated by the larger value bar for agrees and the large bar for strongly agrees. When compared to questions 1-20, the AR count for questions 21-30 was statistically significantly higher with a p-value of 2.9e-5.

**V. CONCLUSIONS AND FUTURE WORK**

Students in the second semester of introductory physics courses at Siena College completed activities using the MARVLS App where they explored augmented reality models of abstract and 3D concepts that students find difficult. The results of the study indicate that the students engage with the AR models and capture different aspects of the models with their drawings. For the assessment questions on magnetism, the students report that the augmented reality models influenced their answer choices. Gains on the EMCA assessment were modest. The calculus students improved their spatial visualization skills as demonstrated in the PVRT results in Table II. For the calculus students we see evidence for transfer in rotational skills through the use of the MARVLS models, while the difference in PVRT performance for life sciences students warrants further investigation. A speculative reason

![FIG. 5. Likert scale results for students indicating that the AR activity influenced their answer choice for question 21 of the EMCA.](image)

<table>
<thead>
<tr>
<th>Question</th>
<th>Avg Pre</th>
<th>Avg Post</th>
<th>AR Pre</th>
<th>AR Post</th>
<th>AR Likert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q21</td>
<td>8</td>
<td>21</td>
<td>0.40</td>
<td>0.44</td>
<td>30</td>
</tr>
<tr>
<td>Q22</td>
<td>12</td>
<td>17</td>
<td>0.33</td>
<td>0.38</td>
<td>30</td>
</tr>
<tr>
<td>Q23</td>
<td>19</td>
<td>36</td>
<td>0.69</td>
<td>0.67</td>
<td>32</td>
</tr>
<tr>
<td>Q24</td>
<td>5</td>
<td>17</td>
<td>0.33</td>
<td>0.39</td>
<td>32</td>
</tr>
<tr>
<td>Q25</td>
<td>9</td>
<td>7</td>
<td>0.14</td>
<td>0.17</td>
<td>26</td>
</tr>
<tr>
<td>Q26</td>
<td>7</td>
<td>28</td>
<td>0.54</td>
<td>0.44</td>
<td>27</td>
</tr>
<tr>
<td>Q27</td>
<td>15</td>
<td>8</td>
<td>0.15</td>
<td>0.26</td>
<td>23</td>
</tr>
<tr>
<td>Q28</td>
<td>5</td>
<td>21</td>
<td>0.40</td>
<td>0.42</td>
<td>27</td>
</tr>
<tr>
<td>Q29</td>
<td>23</td>
<td>21</td>
<td>0.58</td>
<td>0.56</td>
<td>26</td>
</tr>
<tr>
<td>Q30</td>
<td>14</td>
<td>21</td>
<td>0.40</td>
<td>0.33</td>
<td>26</td>
</tr>
</tbody>
</table>

**TABLE IV.** Average pretest and post-test EMCA [18, 22] scores, item difficulty index for this study and a prior study [17], and the results of the AR Likert scale values.

is the tendency for life science students to utilize rote memorization instead of applying a process. To continue this work, we will modify the augmented reality activities to highlight the concepts in the model students may have missed without intentional direction. We will also review student work on the activities along with EMCA scores to understand whether their drawings and answers on the activities correlate with their assessment results. We are conducting interviews with students to understand whether the students believe they benefit from the manipulable AR visualizations. We intend to bring the MARVLS App to a larger audience of colleges and universities and run additional studies to further develop the app and to study the impact of these visualizations on student learning of 3D and abstract concepts in physics. Larger samples will allow for deeper comparisons.

**ACKNOWLEDGMENTS**

The authors acknowledge support from the National Science Foundation award 2120446.
[10] McColgan, M., Hassel, G., Conference presentation for the Technologies Session: Developing Augmented Reality Modules to Teach Physics, 2022 AAPT Summer Meeting, Grand Rapids, MI.
[16] https://mergeedu.com/cube
[22] The selected EMCA questions involve: (Q21, Q28, Q29) magnetic field of current-carrying wires, (Q22, Q30) force on a moving charge in a magnetic field, (Q23) magnetic field of bar magnet poles, (Q24) change in magnetic flux / induced current, (Q25) motional EMF for a bar, (Q26) magnetic force on a current-carrying wire, (Q27) magnetic flux, or see [18] for more details.
"Academia, as a whole, is structured entirely without any consideration for neurodivergency," and other things neurodivergent students want you to know

Liam G. McDermott and Nazeer A. Mosley
Physics and Astronomy Department, Rutgers University, 136 Frelinghuysen Rd, Piscataway, NJ 08854

As more neurodivergent and Neuroqueer students are entering higher education, it is important for educators to be cognizant of and understand what it means to be neurodivergent. Importantly, we must uphold and give space to neurodivergent folks’ narratives to best understand their experiences and identity, and for us to best engage and support our students. While neurodivergent folks are overrepresented in STEM, there is very little research examining the phenomenon of being neurodivergent in the STEM classroom. As a part of a larger study into this phenomenon, we asked neurodivergent folks at various stages of their careers in physics the question: "is there anything you wish professors or colleagues knew about being neurodivergent?" We here report on specifically neurodivergent students’ (undergraduate, graduate, and post-baccalaureate non-academic) responses, and discuss the results drawn from them. We then follow this discussion with recommendations for praxis.
I. INTRODUCTION

Neurodivergent folks are overrepresented in STEM fields [1, 2]. However, when it comes to total enrollment in higher education and eventual completion of their degrees, studies show that neurodivergent students are far less successful than their neurotypical peers [1, 3, 33]. This is especially true for folks identifying with multiple disabilities (ie. being both autistic and dyslexic)[3]. This leads us to an unfortunate conundrum as a field: despite gravitating towards STEM, and being otherwise academically capable [33], STEM fields are regularly inhospitable to neurodivergent students. Dwyer et al. [33] put forth some reasons as to why this is, including stigma based on disability status [5–8], differences in communication styles [9, 10], and intersecting academic/non-academic disability-related barriers [11, 12], and the literature backs these claims up. Still, there is much work to be done to ascertain the full picture [13], and understanding how academia systematically marginalizes and neurodivergent folks is but one small piece to the puzzle. This paper serves as an additional viewpoint, in conversation with the literature, to understand the experience of neurodivergent students, and fill some gaps in what educators should know about being neurodivergent.

There has been much progress in academia with regard to including the voices of disabled people in research [14, 15], and we intend to continue this progress here. Echoing the disability rights mantra "nothing about us, without us," [16] we use this paper to elevate the views and experiences of neurodivergent students to answer the question: what should neurotypical educators know about being neurodivergent? To do this, we report on 11 physics students’ answer to this question, and qualitatively analyze their responses. These students, whom we will call Wren, Elizabeth, Dmitri, Roger, Tom, Quincy, Sky, Catalina, Henry, Albert, and Louis, each provide an interesting viewpoint on both their reflexive identity and beliefs about being neurodivergent, as well as their relational identity and meta-cognition about how others perceive their neurodivergent identity. In this paper, we will narrativize these responses to form a collective idea on what educators should be cognizant of when it comes to their students. In doing so, we synthesize these responses into four main points. Due to space constraints, we will not be developing individual cases for each subject, however we will attribute ideas to each person as we discuss them, and include quotes as examples of each point.

II. POSITIONALITY

It is important when conducting qualitative research to acknowledge how one’s identity and experiences influence the analysis of data [17, 18]. It is also critical to the spirit of this research that analysis is done with members of the neurodivergent community instead of simply about them. To this effect, we note that author McDermott is a neurodivergent physics education researcher. He is also multiply disabled, being hard-of-hearing as well. Author Mosley does not identify as disabled. Both authors identify within the LGBTQIA+ community. We also note that author McDermott identifies as white, while author Mosley identifies as Black. Each author brings an interesting perspective to this research, and we found our respective identities to compliment each other, with our life experiences helping to check each other’s biases, prejudices, and neurotypical-normative expectations for how minds/neurotypes should work. Disabled folks and even the most well-intentioned abled folks can, and do, commit ableist acts and analyses [19, 20], so it is critical that these are checked by teammates.

Author McDermott being multiply disabled and neurodivergent allows him to take on a role as an insider to the neurodivergent community. In terms of data collection, this looks like quicker establishment of rapport, and a unique understanding of how to best navigate and structure questions on the fly that are understandable and relevant to a subject during a semi-structured interview. In analysis, being neurodivergent helps understand community-specific subtext which would otherwise be missed by a neurotypical person. In a similar vein, author Mosley’s Black identity allows them to take a non-white-centric approach to data analysis, providing insight into oppressive structures, of which McDermott’s privileged standpoint obscures from his view.

III. NEURODIVERSITY, NEUROQUEER, AND THE NEURODIVERSITY APPROACHES

Neurodiversity is a concept developed in 1998 by autistic scholar and activist Judy Singer [21]. It is, put simply, a celebration of the diversity of minds. Neurodiversity, initially situated as autistic self-advocacy grown from challenges to the medical model of disability and resistance to disability as a means of social control, soon grew to embrace non-autistic "cousins" (folks identifying as dyslexic, dyscalculic, ADHD, etc) [22, 23], and eventually more "relatives" (folks identifying as epileptic, OCD, schizophrenic, Tourette’s, etc)[23, 24].

Further work has since reinforced concepts central to Singer’s original neurodiversity concept, such as the anti-capitalist philosophy which rejects exclusion of disabled folks from society on the grounds of "productivity" [21, 25, 26], and the rejection of the socially constructed "normal" [21, 25, 27]. Scholars have since built on these concepts, introducing ideas from Crip Theory [28, 29] and Gender and Sexuality Studies [30, 31] to assert neurodiversity as a Queer thing: a Neuroqueer thing [30, 31]. As such, being neurodivergent and claiming neurodivergency as an identity serves as a tool (in Lorde’s [32] "Master’s Tools" sense) to question, subvert, and dismantle neurotypical and ableist norms.

For research and praxis involving neurodiversity, Dwyer [33] discusses the “Neurodiversity Approach” as a way to model disability, stating "Instead of there being one singular neurodiversity paradigm, it might be more accurate to
speak of multiple 'neurodiversity approaches'" (p. 75). As Singer [21] asserts in her seminal work, neurodiversity is grounded in direct opposition to the medical model of disability, which describes disability as something negative which is to be cured to achieve social equality for disabled people. Traditionally, when eschewing the medical model, disability scholars turn to the social model of disability, which describes disability as socially constructed based on neutral natural human variation, and thus social equality is achieved through dismantling barriers to access. However, this binary, Dwyer and Singer assert, is counterproductive to social change and does not reflect the reality of neurodivergent identity [21, 22, 33]. Especially when considering disability situated within the mind, binaries do not sit well. Neurodiversity, instead of existing in the social/medical binary, sits somewhere between. Being neurodivergent is interactional. It’s relational based on a person’s inherent qualities (like the medical model [34]) interacting with their environment and vice versa (like the social model [34]). Critically, the neurodiversity approaches and research/praxis which include them reject the idea of "curing" neurodiversity, and instead assert that neurodiversity should be a celebrated part of the human experience.

No movement is above critique, and we would be remiss if we did not discuss critiques of the neurodiversity movement in this paper. Critics of neurodiversity, the movement, the paradigm, and the approaches, state that the theory centers "high functioning" neurodivergent folks, and invalidates/silences those who are nonverbal, "low functioning," etc. [35, 36]. This concern is valid, and research which centers neurodivergent people must be aware of this and do everything possible to uplift all voices of neurodivergent people [15, 38, 39]. One push-back to this, however, lies in this dichotomy of high/low functioning, the neurodiversity movement has space for all [37, 40]. Core to the neurodiversity movement is the idea that high and low functioning labels oversimplify neurodiversity as everyone has both low and high needs to exist in the world [41]. Instead, there is a spectrum of folks whose maintenance of their bodies and minds are considered too much for a neurotypical-normative or abled-normative society.

Other critics indicate concerns about how the neurodiversity movement is made up of a majority white population, who are often of privileged class [35, 42]. These are extremely valid concerns, and there are many explanations as to why this is. White people are more likely to be believed by doctors and receive diagnoses in line with being neurodivergent, and consequently the care needed to set them on the course to feeling they belong in the neurodiversity movement [43]. In addition, people who are from lower class backgrounds, especially in the USA, often do not receive medical care, including mental and psychological healthcare [44]. It is therefore important that we take an intersectional approach to this and future research on neurodivergent experiences, narratives, and identities, and that we do what we can to assure a representative sample in our research [45].

IV. DATA COLLECTION AND METHODOLOGY

As part of author McDermott’s dissertation study examining the phenomenon of what it means to be neurodivergent in physics, he collected interviews with 19 neurodivergent physicists at various stages of their careers. This includes undergraduates in at least their 3rd year of school, graduate students, post-baccalaureate non-academic physicists, and post-graduate physicists. Subjects were collected using a convenience sampling method, based on subject relationship to author McDermott. Subjects took part in a 1-2 hour semi-structured interview in which they were asked questions pertaining to their experience as physicists, as neurodivergent people, and the intersection of those two identities. Of these 19 physicists, we here report on only 11 students’ responses.

Subjects were asked, as the final question in the interview, “is there anything you wish professors or colleagues knew about being neurodivergent?” This question was initially included to assure that we hadn’t missed any information about being a neurodivergent physicist, as this is a phenomenology study, and as a friendly way to reinforce the guiding principle of this project: that we wholeheartedly and enthusiastically want to include the interviewees as active participants in the research process. Starting from our first interviewee, however, it became evident that this question warranted special attention. This question, coming after a lot of structure to get at specific facets of identity, allowed for an unstructured re-prieve for folks to speak their mind on subjects near and dear to the heart of their identity.

On the subject of inclusion of subjects as active participants in the research process, we engaged in member checking. Subjects were asked to review this paper before submission and provide any alterations they felt was necessary to best reflect their views accurately and responsibly. We have no reason to believe subjects would agree on all points to appease the researchers, as they have no incentive to do so, and were repeatedly assured throughout the interview that they are in full control of their own information.

Subjects were asked to self-identify at the end of the interview. Of the 11 subjects, 4 identified as men, 5 as women, and 2 as non-binary. 4 subjects identified as trans/nonbinary, and 7 identified as cisgender. 3 identified as post-baccalaureate non-academic physicists, 2 as graduate students, and 6 as undergraduate students. All 11 subjects identified as white.

We analyzed the data using thematic analysis. We took a latent approach to the analysis, focusing on subtext and using experience gained through being in the neurodivergent community. We took an inductive approach to this method, individually familiarizing ourselves with the data, coding the data, and coming together to discuss and find consensus in coding. There were no instances where we were not able to reach consensus. We then organized our codes into four themes, discussed in our Results section. After writing-up this paper, themes were verified by subjects via our member checking process to ensure proper representation of their ideas.
V. LIMITATIONS

No study is without limitations, and this study is certainly no exception. The most glaring example of this to the authors is that the entire subject population is white. This means that our data reflects a very white view of neurodiversity. Neurodivergent identity as it intersects with race is a topic of much valuable discussion [42], and we are going to remedy the lack of a racially diverse subject population in further reports using an expanded dataset. Another limitation to this study is that we are specifically examining only "success stories" of neurodivergent physicists (those who stay in physics or retain their identity as a physicist). In limiting the data, however, we gain insight into what works in driving student success in addition to what does not.

VI. RESULTS

One thing that must be noted here is that the neurodivergent community is a remarkably heterogeneous one [46]. There is no such thing as the one way to be neurodivergent. That being said, we are a community of people who have come together due to similar experiences being marginalized by society due to differences in how our minds work. The responses to the question "is there anything you wish professors or colleagues knew about being neurodivergent?" are exemplary of this idea. No two people presented in this paper have the same identity, and none of them have the same experiences. Students we interviewed identified with Obsessive Compulsive Disorder (OCD), Dyslexia, Attention Deficit Hyperactivity Disorder (ADHD), autism, epilepsy, Bipolar Depressive Disorder (BPD), and Schizoaffective Disorder. Students were cisgender, transgender, and nonbinary. They also identified as bisexual, heterosexual, and demisexual. However despite this heterogeneity, it was remarkably easy to condense the 11 responses to four themes. Below is this list of things that neurodivergent students want you to know:

- Professors need to both understand and accept that they are ill-equipped to teach neurodivergent students.
- Coming out as neurodivergent is an often last-ditch effort to succeed in class due to stigma, and is a very vulnerable thing.
- Being neurodivergent is a fundamental aspect of being oneself.
- Not everyone knows they are neurodivergent.

We will break down each point in the next four paragraphs, though it should be noted that some quotes may reify multiple themes at once.

Regarding ill-equipped professors: The most common thread which linked the responses is that, in these neurodivergent students' experience, professors and those in power are ill equipped to teach neurodivergent students. To this effect, the only thing Quincy cites as what she wanted professors to know, "I just wish [professors] as a whole were more informed." Wren, perhaps, put it best, giving us the title: "I wish there was an understanding that academia, as a whole, is structured entirely without any consideration for neurodiversity." This sentiment is echoed by Henry, who states, "we need to be educated differently. It doesn't mean we're stupid or can't learn, it's that we can't learn and truly thrive in the current state of physics education." Folks also acknowledged that, for right now, we don't have all the answers, but students understand that there is a fundamental problem with how we teach neurodivergent students. This idea is held up by Albert, who says "I'm trying to figure out myself... I have still not figured out exactly what I'm working with here." That is not to say that there isn't insight to be gleaned, however. Tom pointedly states, "neurodivergence really can turn up [little] differences... and sometimes that makes people very difficult to work with. But patience is probably the best medicine."

Regarding coming out as neurodivergent: Elizabeth gives an especially powerful statement on our second point. "I wish sometimes that [professors] knew that I'm not trying to be neurodivergent. The last thing I want to do is to have to ask for accommodations." She expands this statement outwards, "for a lot of people it's very difficult to even start that conversation. So I think if you get to a point where your student is coming up and asking you 'hey, I need this accommodation,' think how much of an internal struggle that student had to actually get to this point. And you should acknowledge and be cognizant of that." It's important also to acknowledge that students who ask for accommodations at the beginning of the semester have likely had to go through the process of struggling to receive accommodations from other professors many, many times before. The fact that they are approaching a professor with this request is because they know that this is the solution that works for them. It is also incredibly important to understand, especially if professors have never had to go through this process personally, that accommodation requests phrased in the form of a question should not be treated as requests, and are usually phrased as questions for the sake of politeness, as English speakers often use conventionally indirect strategies to assert need [47]. When asking for accommodations, students are not asking, and it is counterproductive to their learning to treat it as such. Sky puts it well, saying "if neurodivergent people are telling you that they're neurodivergent in whatever way that is, that it's them being vulnerable and honest and it's not something to look at as an excuse."

Regarding the fundamentality of being neurodivergent: while it is possible to hide being neurodivergent for some, hiding it (masking) takes a mental toll [48], but revealing it to others opens oneself up to ridicule [49], or worse physical or social violence [50]. Henry speaks to this: "we are fundamentally different and think differently than a neurotypical person, which means we need to be educated differently." Louis seconds this, saying, "the way in which we learn is not always the same as a non-neurodivergent [person]." Students, critically, want professors to know that the mismatch between the fundamentality of being neurodivergent and how the rules of the classroom are built is why they may miss class, not turn
in work, or any number of things which cause their grades (and potentially their relationship with professors) to falter, not that they simply don’t care. Roger implores, "I wish they knew that even if, like, when I was like really struggling and not train homework and not going to class... it wasn’t because I hated the course or anything like that, it was because of my neurodivergence," This fundamentalism is, critically, not to be viewed as deficit, but to be viewed as an important and good part of the human experience. Sky shares her perspective on this. "For one thing, I feel like the more we learn about neurodivergency, and the more we learn how many people there are who are neurodivergent and how many types there are, and how truly incredible the brain is of coming up with different ways to maneuver, to craft itself." She concludes, "I think it’s something to be celebrated."

Regarding many people not knowing they are neurodivergent: Tom speaks to our fourth point, saying, "it’s a complex and unique experience for each student. Everyone’s a little bit different... and sometimes people are not aware of that in themselves." Multiple subjects cited that having empathy towards neurodivergent students not only helps them succeed, but helps us be kinder to ourselves, even if we’re assigned neurotypical. In fact, many subjects spoke about their own journeys towards being cognizant of their neurodivergent identity, often because others helped them succeed because of their differences, not in spite of them. Louis states, regarding "I guess it’s not with every professor. Again you’ll have professors who are quite open about this especially. Neurodivergent professors, as you might expect." Catalina furthers this idea, saying, "I feel like there’s a lot of [professors] that are neurodivergent, but don’t really acknowledge it, or don’t really know it, or they just came up with really good coping mechanisms." She continues, "That’s something that should be acknowledged, there’s a lot more people out there that are neurodivergent than those who are willing to acknowledge.” Many people simply don’t know they’re neurodivergent due to differential access to healthcare or being in an environment which so heavily encouraged masking (neurodivergent folks covering up being neurodivergent by acting neurotypical). It is therefore critical, not just to folks who are comfortable in their identity as neurodivergent, but also for folks who have not made that journey yet, that we pay special attention to how we teach our students and that we teach in a neurodiversity-inclusive way.

VII. RECOMMENDATIONS FOR PRAXIS

Where do we go from here? It is clearly not enough, from our first point, to simply point out that a problem exists in educating and including neurodivergent students in higher education. Action must be taken to educate educators on good changes to make to curriculum and classroom/laboratory practices. Even more importantly, action must be taken by educators to enact good change. The authors said this before in previous work [24, 25, 51–53], and other Physics Education Research scholars concur [19] and we will say it again: ableism is done by good people. It is done by good actors who believe in equity and fairness, and is often done unwittingly or in the name of pragmatism. Ableism is normal in academia [54], and it is structural through and through. It takes good people, good allies resisting this system of oppression, being vigilant of ways that violence is enacted, often invisibly [55], toward disabled people. We therefore provide the following framework of six tasks for changing praxis to create an neurodivergent-inclusive classroom and laboratory:

1. When possible, be flexible with documentation requirements. Not everyone can afford to see a doctor, and many learn what access needs they require through trial and error. Further, do not treat accommodation requests as questions. They are demands.
2. Incorporate pedagogical practices inclusive of neurodivergent students’ access needs, such as Universal Design for Learning.
3. Include the narratives of disabled people in your lessons and storying of physics.
4. Include neurodivergent students in your research.
5. Be flexible with due dates and mandatory attendance.
6. Assume that students are, first and foremost, competent at the tasks they are given, and work up from there to help their performance align with their competence.

Much like Amendment IX of the United States’ constitution, we acknowledge that we have not given every possible recommendation for praxis which ensures the inclusion of neurodivergent students in the physics classroom. We are also aware that certain institutions have very strict processes for students and professors to follow when engaging with accommodation requests. An incredible way to dismantle oppressive structures is to act within them when we have the privilege to do so, but to always remember Lorde’s words, that "the master’s tools will never dismantle the master’s house" [32]. It is critical that those with privilege not sit idly by and simply be nice able-bodied people. It is important to remember that social acts are shared acts, communication especially [20]. Additionally, we must remember that being disabled and being neurodivergent are both remarkably individual things, and thus what works for one neurodivergent person may not work for every neurodivergent person. It is therefore important to engage in productive two-way communication with students. If you see students are struggling in class, reach out to them and ask to work together to find solutions. When students reach out to you, work together with them to find solutions. Be kind above all else.

ACKNOWLEDGMENTS

Special thanks to Geraldine L. Cochran for her work as advisor for this study. This work was funded by LGBT Tech’s PATHS STEAM Education Grant, and Rutgers University Physics and Astronomy Department.


Evaluating learning of motion graphs with a LiDAR-based smartphone application

Colleen Megowan-Romanowicz (she/her)
American Modeling Teachers Association, Sacramento, California, USA

Daniel J. O’Brien (he/him)
Department of Physics, Georgetown University, Washington, D.C., USA

Rebecca E. Vieyra (she/her), Chrystian Vieyra Cortés (he/him)
Vieyra Software, Washington, D.C., USA

Mina C. Johnson-Glenberg (she/her)
Department of Psychology, Arizona State University, Tempe, AZ, USA

Data modeling and graphing skill sets are foundational to science learning and careers, yet students regularly struggle to master these basic competencies. Further, although educational researchers have uncovered numerous approaches to support sense-making with mathematical models of motion, teachers sometimes struggle to enact them due to a variety of reasons, including limited time and materials for lab-based teaching opportunities and a lack of awareness of student learning difficulties. In this paper, we introduce a free smartphone application that uses LiDAR data to support motion-based physics learning with an emphasis on graphing and mathematical modeling. We tested the embodied technology, called LiDAR Motion, with 106 students in a non-major, undergraduate physics classroom at a mid-sized, private university on the U.S. East Coast. In identical learning assessments issued both before and after the study, the mean score of students working with LiDAR Motion improved by more than that of those using standard-issue sonic rangers. Further, per a voluntary survey, students who used both technologies expressed a preference for LiDAR Motion. This mobile application holds potential for improving student learning in the classroom, at home, and in alternative learning environments.
I. INTRODUCTION

Although time-based motion graphs are fundamental to teaching and learning kinematics, students’ intuitions about how to interpret them are often misaligned with what the graphs actually represent [1]. That students struggle to read graphs has been well-documented since the earlier days of physics education research [2–5]. The resulting taxonomies of student difficulties on this topic became the basis for heavily used concept inventories, such as the Mechanics Baseline Test [6] or the Test of Understanding Graphs in Kinematics (“TUG-K”) [7].

Educators have proposed several pedagogical and technological supports to help students make sense of motion graphs. Early on, educators conceived of microcomputer laboratories with commercially-designed, specialized ultrasonic rangers attached to a computer visualization output [8–10]. Seeking other technologies that are more accessible, versatile, or cheaper, recent proposals have incorporated ultrasonic rangers on Arduino boards [11], robots [12, 13], and Kinect for Xbox One [14]. Despite smartphones’ potential for collecting and visualizing physics-related data for many topics [15], the low-cost mobile accelerometers embedded within them are generally not adequate for capturing accurate position data—at least in introductory physics contexts—due to a need for regular recalibration and data cleaning [16].

The addition of light detection and ranging (LiDAR) technology to a number of iPad and iPhone models since 2020 provides a radical new opportunity to measure motion with high-precision position sensors. Through our creation of a new app, iPad and iPhone users now have the capability to visualize graphs of their own motion in real time, providing new opportunities to understand kinematics using technology that many learners already have in their pockets. Research supports that when the immediacy of motion visualization is paired with a learner’s own motion there are significantly positive learning outcomes [17]. This paper presents the outcome of an introductory-level-university-physics study using the LiDAR Motion app (www.lidar-motion.net) for LiDAR-enabled iPads and iPhones. It demonstrates the utility of this new technology, especially considering many teachers’ need to flexibly respond to circumstances such as distance learning following the COVID pandemic [18].

II. TECHNOLOGY FOR QUANTIFYING LINEAR MOTION

This study tested the usefulness of LiDAR Motion, a free mode within the Physics Toolbox Sensor Suite application available for iPad and iPhone. It was developed to plot a user’s movement as they move directly perpendicular to a flat wall or object (Fig. 1). The LiDAR hardware in the mobile device emits an array of infrared laser pulses and then senses the reflected pulses, using the light’s time of flight to calculate the distance from the phone to the object it is pointing toward. The resulting information is used by the app to display real-time data on position-time and/or velocity-time graphs on the screen (Fig. 2), allowing users to understand how the graphs coordinate with their movements. Additionally, the app includes a game challenge feature in which users must attempt to match their motion to pre-drawn graphs. Each challenge was developed to address common student difficulties, and users are only successful once their motion data is validated against expected values within an acceptable $R^2$ threshold. When the user has moved in a manner that stays within a prescribed error zone around the displayed graph, the app displays a congratulatory “Great job!” message with confetti.

III. RESEARCH QUESTIONS

Considering the increasing access that teachers and learners have to mobile devices, the purpose of this study was to determine whether physics students learn about
motion differently when using the LiDAR Motion application than when using specialized commercial science equipment: an ultrasonic ranger, track, cart, and computer graphical output. We especially hoped to gauge whether LiDAR Motion helps students better model motion and interpret kinematics graphs. Understanding the affordances of these technologies within the same inquiry-based pedagogical approach will shine a light on how teachers might prioritize technology selection for teaching motion graphs.

IV. METHOD

The technologies were tested in a non-major, introductory physics course at a mid-sized, private university on the U.S. East Coast. Eight laboratory sections were assigned to either the intervention or the control: four sections were assigned to use LiDAR Motion and four sections were assigned to use Vernier ultrasonic rangers paired with LoggerPro. Sections were distributed such that sections of a particular characteristic (e.g., time of day or class size) were not consistently assigned to one technology. Regardless of technology, students were placed into pairs and asked to take turns as they carried out laboratory tasks.

The laboratory sections were led by the course’s single regular instructor and supported by at least two undergraduate or graduate teaching assistants, all of whom were independent of this project’s research team. Additionally, one of the researchers met with the course instructor and the teaching assistants a week prior to the laboratory sessions to introduce them to the phone application and to the Vernier equipment and software. The instructor and assistants were briefed on the type of interaction and feedback that should be offered to students during the sessions. At least one researcher was present during each of the laboratory sessions to observe student interactions.

Students’ understanding of kinematic concepts was evaluated quantitatively using a written assessment. The assessment was composed of fifteen questions about position and velocity graphs sampled from the Test of Understanding Graphs in Kinematics v4.0 [19] and Diagnoser [20]. The test was administered to students one week prior to the laboratory session (pre-test) and immediately following the laboratory session (post-test). The pre- and post-test results were compared against each other and between the intervention and control groups.

Students in each section were provided with nearly identical protocol sheets to guide them through the lab session (available at www.lidar-motion.net/lesson-ideas). The only differences pertained to specific instructions about the technology they were using. The protocol sheets began with a free play session during which the students could explore the technology to which they were assigned. Subsequently, students were guided into a set of tasks to challenge their understanding of kinematics and motion mapping. This feature is built into LiDAR Motion as a graph-

FIG. 3. A student walks toward and away from an ultrasonic detector while her lab partners observe her motion on a graph.

matching game in which the user walks perpendicular to a wall to match the shape of a position-time curve to their movement. Students using the ultrasonic detector were instructed to use their bodies to walk in front of the detectors (Fig. 3) to create position-time motion graphs drawn on the protocol sheet—which mirrored those in the LiDAR Motion graph match game. All sections completed the protocol within 45-60 minutes, regardless of the technology they had been assigned.

Lastly, one separate laboratory section employed both forms of technology, with half of the students starting with LiDAR Motion, the other half of the students starting with ultrasonic rangers, and both groups switching halfway through the tasks. Students in this section were also asked to voluntarily complete an interest form after the laboratory session. This form gauged students’ impressions of the level of enjoyment of the activities, the level of usefulness of the activities to learning the concepts, the level of engagement of the activities, and which technology they preferred using: LiDAR Motion or sonic rangers. The form also had a free-response prompt for students to comment on the activities.

IV. RESULTS

The written assessment test was used to gauge the effectiveness of LiDAR Motion relative to standard classroom technology for the eight sections that completed the full protocol with a single technology. The maximum score on the test was 15, and pre-test scores in all sections were relatively high (Fig. 4a). The two groups (LiDAR and ultrasonic) were matched at pretest, t(104) = 1.33, p = 0.19. By post-test, the LiDAR Motion group increased by a statistically significant 1.32 points, t(56) = 4.12, p < 0.001, and the ultrasonic group improved by a statistically significant 0.53 points, t(48) = 2.65, p < 0.05. Per an ANOVA comparing these gains (post - pretest), the increase in the LiDAR group’s mean score after intervention was greater than the increase in the sonic group’s mean score at a statistically significant level, F(1, 104) = 4.02, p < 0.05.

The number of students who answered correctly on each assessment question was also compared. Students in the LiDAR Motion sections improved more on twelve of the fifteen questions (Fig. 4b). Raw data illustrating the
Students in the ninth laboratory section who completed tasks with both technologies were asked which technology they preferred. Of the fourteen individuals who responded, eleven students preferred LiDAR Motion, two preferred the ultrasonic ranger, and one had no preference (Fig. 5a). Those who preferred LiDAR Motion rated the laboratory activities’ level of enjoyment and engagement more favorably on a scale from 1 to 5 (Fig. 5b). However, given the small number of students who preferred the ultrasonic ranger or had no preference, inferential statistics were not calculated.

Qualitatively, researchers observed the nature of student interactions with the technology during the laboratory sessions. Students were on task the majority of the time in both the treatment and control laboratory sessions, and all pairs were highly collaborative. However, those using the app expressed more frequent and more intense emotions, including fist bumps, high fives, and exclamations upon achieving a graph match. One student shared with her lab partner, “It’s actually kind of fun. I wish I had this in high school.” Many students also expressed a sense of anxiousness about doing the graph match tasks: “My heart is beating so fast,” and “It makes me so nervous. It’s a good nervous...[a] pressure to get it right.” Some students expressed frustration about not being able to match the graph: “I know what I’m supposed to be doing, but my mind and my feet aren’t coordinating.”

Although tasks in the LiDAR and ultrasonic ranger protocols were designed to mirror each other—requiring students to walk linearly to replicate a particular graph shape—students sometimes deviated from the written protocol. For example, multiple groups using ultrasonic rangers instead used their arms to move a book along a track in front of the detector to make graphs, detracting from the intended locomotive embodiment of the activity (Fig. 6). Moving the arm is considered a lower level of embodiment compared to walking, which activates more sensori-motor neurons, among other systems [21]. The real-time response and handheld nature of the iPhone app kept students from deviating in a similar way from the protocol while using LiDAR Motion, though at least one student was seen moving the phone towards and away from the wall by extending and retracting their arms, rather than by walking.

IV. DISCUSSION AND CONCLUSION

This research study illustrates that LiDAR Motion results in more significant achievement gains for
undergraduate non-major students when compared to those who used commercial ultrasonic rangers and associated software. The question-by-question results offered a more granular perspective on learning gains. For example, in contrast to those using ultrasonic rangers, students using LiDAR Motion improved by the greatest margin on assessment questions 1, 2, 6, and 7; these four questions all relate to single-segment analyses of constant velocity. Although one can attempt to match a graph on LoggerPro, the user is not given real-time feedback on accuracy. The feedback on graph-match goodness of fit in LiDAR Motion may contribute more to students' comprehension of motion, especially for simpler graphs. The app forces students to re-evaluate and re-attempt their movements until it validates their graph match with positive feedback. Matching graphs using the whole body and doing that task multiple times until "correct" may have led to the increase in gain scores for the LiDAR group.

Contrary to the belief held by some that smartphones may pose a distraction in learning environments, our own study and related studies by Kaps et al. [22] and Mazzella and Testa [23] showed that mobile devices can be highly effective teaching tools. Further, we observed a strong preference for the use of LiDAR Motion, which echoes findings by Hochberg et al. [24] and Ozkan [25] that students' enjoyment, interest, and curiosity in science increase when using mobile devices.

One notable aspect of this study is that students displayed highly collaborative engagement, with or without the use of the smartphone. Prior to the study, the researchers had some concerns that the use of a personal mobile device, with its small screen, might threaten students' cooperative actions. However, these concerns were unfounded, as students in the intervention and control groups could be seen discussing their planned motions and coaching each other. This finding stands in contrast to Anderson and Wall [14], in which the use of Xbox One decreased student collaboration—although this was potentially a result of the fact that they used a single piece of technology with the entire class, while the present study had one set of technology for each pair of students.

Our findings are limited by a number of factors that can be mostly attributed to the advanced nature of student understanding. Although the research participants were introductory-level, non-major physics students, the highly competitive nature of the university suggests that many of these students had prior physics classes in high school and had already developed good graph reading skills. Some students also mentioned that they had previously used sonic rangers. The effect was that students had substantial prior knowledge, and pre-test scores did not reflect an expected normal distribution. Future studies should consider working with a more general audience or students much earlier in their learning trajectory, perhaps including secondary students who are just learning algebra (around ages 12-14). Alternatively, more difficult assessment questions could provide a deeper understanding of upper-level students' learning gains with the technology.

Availability of the LiDAR Motion tool is currently limited by the number of smartphone models with LiDAR hardware. At present, direct-time-of-flight LiDAR sensors are available on the Pro and Pro Max models of iPhone 12, 13, and 14, as well as the iPad Pro models released from 2020 to present.

The present study is particularly applicable in schools that have policies that include iPads or permit the use of personal devices, but may not have the budget for specialized science equipment. It is also applicable in cases where students need to learn from home. Our study places tools such as iPads and iPhones with LiDAR Motion among the set of viable options from which educators can select to effectively teach about motion graphs.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation Grant #2114586 and approved by Georgetown University Institutional Review Board (Study00005454). Special thanks to Christopher Cothran and the physics teaching assistants of Georgetown University. Special appreciation to Thérèse Vieyra, who assisted in the development and testing of this activity.


How traditional physics coursework limits problem-solving opportunities

Barron J. Montgomery and Argenta M. Price

*Physics Department, Stanford University, 450 Serra Mall, Stanford, CA, 94305*

Carl E. Wieman

*Physics Department and Graduate School of Education, Stanford University, 382 Via Pueblo, Stanford, CA, 94305*

A major goal of physics education is for students to develop problem-solving skills. To become expert problem-solvers, students need to deliberately practice those skills. In this analysis, we defined problem-solving skills as a set of 29 decisions that were previously identified as defining the problem-solving process of expert scientists. We quantified the amount of practice undergraduate physics students get at making each decision by coding the decisions required in assignments from introductory, intermediate, and advanced physics courses at a prestigious university. A research-focused capstone course was the only example that offered substantial practice at a large range of decisions. Problems assigned in the traditional coursework required only a few decisions and routinely removed potential opportunities for students to make other decisions. This analysis suggests that to better prepare undergraduates for solving problems in the real world, we must offer more opportunities for students to make and act on problem-solving decisions.
TABLE I. Problem-solving decisions characterized by Price et al. Note: numbers are for reference, not meant to imply a sequential order of decisions.

<table>
<thead>
<tr>
<th>Decision description</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
</tr>
<tr>
<td>D2</td>
</tr>
<tr>
<td>D3</td>
</tr>
<tr>
<td>D4</td>
</tr>
<tr>
<td>D5</td>
</tr>
<tr>
<td>D6</td>
</tr>
<tr>
<td>D7</td>
</tr>
<tr>
<td>D8</td>
</tr>
<tr>
<td>D9</td>
</tr>
<tr>
<td>D10</td>
</tr>
<tr>
<td>D11</td>
</tr>
<tr>
<td>D12</td>
</tr>
<tr>
<td>D13</td>
</tr>
<tr>
<td>D14</td>
</tr>
<tr>
<td>D15</td>
</tr>
<tr>
<td>D16</td>
</tr>
<tr>
<td>D17</td>
</tr>
<tr>
<td>D18</td>
</tr>
<tr>
<td>D19</td>
</tr>
<tr>
<td>D20</td>
</tr>
<tr>
<td>D21</td>
</tr>
<tr>
<td>D22</td>
</tr>
<tr>
<td>D23</td>
</tr>
<tr>
<td>D24</td>
</tr>
<tr>
<td>D25</td>
</tr>
<tr>
<td>D26</td>
</tr>
<tr>
<td>D27</td>
</tr>
<tr>
<td>D28</td>
</tr>
<tr>
<td>D29</td>
</tr>
</tbody>
</table>

This work is related to extensive work on characterization of problems according to various features. Here, instead of classifying problems into categories, we analyze which decisions need to be made as a way of characterizing the process required to solve the problem. Decisions can be applied to other problem characterizations as well. For example, Chi and colleagues [7] found that novices and experts categorize problems differently: according to surface features vs. underlying concepts and solution approaches. This categorization relies on the solver’s schema or mental model of the problem, which is reflected as making D4 (important features) and D5 (mental model) in the set of problem-solving decisions. The characterization of problems as well-defined (or well-structured) and ill-defined is particularly relevant [8]. Problems that would be categorized as ill-defined require solvers to make D3 (goals), and likely D13 (what info need). Ill-defined problems are also described as having multiple
possible solutions and solution paths [9], so would require solvers to make D8 (possible solutions) and D25-26 (reflect on approach, reflect on solution). Problem-solving transfer described by Mayer and Wittrock [10] depends on D7 (related problems). Other work defined a subset of ill-defined problems as complex problems, which are novel, complex, changing, and intransparent [11]. Such complex problems would require most of the problem-solving decisions. Heller et al. [12] have presented "context rich" problems for physics instruction which call on students to make more decisions than in typical physics course problems, and they have outlined a 5 step solution strategy which involves making several of the 29 decisions we present here. Here we are extending previous work to examine exactly which decisions students are being called upon to make in the problems given in actual physics courses.

II. METHODS

We analyzed homework problems from physics courses that are part of the required mechanics sequence for physics majors at a prestigious university: introductory mechanics for freshman potential physics majors ("intro"), junior/senior-level intermediate statistical mechanics ("intermediate"), and a senior-level advanced Lagrangian mechanics ("advanced"). As contrasts to the traditional problems used in these courses, we also analyzed the project from a research-based capstone laboratory course ("capstone"), and homework problems from another university’s introductory mechanics course that were structured around a template designed to require problem-solving decisions ("template"). The intro, intermediate, and template courses were taught in an active-learning style, while the advanced course was a traditional lecture. For consistency of comparison across courses, we analyzed homework problems only, not in-class work. For the capstone course, we analyzed the final research project.

Problem-solving decisions required to solve the homework problems were coded in an iterative process. The research team consisted of a senior undergraduate physics major, a non-physicist discipline-based education researcher with expertise in problem-solving decisions but less physics content knowledge, and a physics and DBER professor who also had expertise in problem-solving decisions. The student researcher solved each problem step-by-step, noting decisions made during the process (for the intermediate and advanced courses, these were solved in real-time while the researcher took the respective courses). For a subset of problems, the student researcher also cross-examined the instructor’s solutions to identify any additional decisions that the instructor intended to be made. The research team then discussed the decisions taken when solving, how the other researchers would have approached the problem, and what phrasing in each problem prompted or removed decisions. We created a code book of definitions of each decision, with physics-problem-specific language and examples. These definitions were iteratively updated after each discussion. The team discussed around 70% of the problems analyzed and reached a consensus on the decisions involved. The remainder (primarily from the intermediate and advanced physics courses) were coded just by the student researcher. We analyzed all homework problems from the intro, intermediate, and advanced courses, but only two problems from the template course because the template structure required the same decisions for all problems. To analyze the research project from the capstone course, we conducted a retrospective cognitive task analysis interview [3], in which the student researcher described their group’s project, explained their problem-solving process, and presented their final report along with some of their analysis work. After the interview, the decisions involved were coded by the interviewer and verified by discussion.

We categorized decisions into two subsets: encountered or removed. Encountered decisions included prompted decisions in which students were explicitly instructed to make a decision as part of their solution; for example, "state any assumptions you need to make." Encountered also included unprompted decisions that students would be required to make on their own to successfully solve the problem. Removed decisions were explicitly made by the problem statement, preventing the student from needing to make that decision. Sometimes two or more decisions would be impossible to differentiate. For example, in the context of these problems D15 (plan) and D16 (calculations) were often the same decision because the planning (D15) required to solve these problems usually involved deciding which equations to use and which mathematical manipulations to apply (D16); in that case, we coded both decisions as required but noted they likely coincided. Figure 1 shows an example of our coding for an intro mechanics problem. We represented this data by calculating the percentage of problems in which each decision was en-
FIG. 2. Percentage of problems in which decisions were encountered (blue) or removed (yellow). All homework problems were analyzed from the traditional physics courses: Intro mechanics (n=26), intermediate statistical mechanics (n=41), and advanced Lagrangian mechanics (n=21). As contrasts, one project from the capstone research-based lab course was analyzed through retrospective think-aloud interview, and two "template" problems were analyzed from a problem-solving oriented intro course.

countered or removed (see Fig. 2). This representation was used to quantify and compare the level of decision-making agency afforded to students for each set of coursework. If a decision was encountered in one part of a problem but removed in a different part of a problem, we counted the decision as being both encountered and removed. If a decision was encountered or removed more than once in a problem, we only counted it once. Of note, for any given problem, many decisions were not given any code - this means the decision was not prompted or explicitly removed and we did not have sufficient information to conclude how likely a student would be to choose to make that decision in solving the problem. Some decisions are easier to infer from a problem statement and student work than others, reflection decisions being particularly challenging unless they are explicitly prompted; this is a limitation of the analysis that future work conducting think aloud interviews should help clarify.

III. RESULTS

We found that the most commonly encountered decisions in undergraduate physics homework were D4 (features), D15 (plan), and D16 (calculations). Decisions D19 (results vs. predictions) and D26 (reflect on solution) were the mostly likely to be explicitly prompted. Meanwhile, the most commonly removed decisions were D10 (simplifications), D11 (decompose), and D15 (plan). When comparing across varying course levels, our data suggest that introductory physics courses eliminate decisions more frequently than intermediate and advanced courses, but none of these traditional courses require many problem-solving decisions.

Our data also reveals a gap in the range of available decisions between the traditional courses and the research-focused capstone course, the latter offering a far wider diversity of decisions for students to make. In addition, we found that problems designed to follow a problem-solving decisions template indeed prompted many more decisions. On average, students only encountered 2.76 decisions per problem in the traditional courses and had 1.6 of decisions removed per problem. All three traditional courses were similar, with the intro course removing the most decisions on average per problem but also encountering the most (Table II). Interestingly, the intro course only required (unprompted) 2.1 decisions on average, which was less than than the intermediate (at 2.3), but included more prompted decisions, for more total encountered.

As noted above, on any given problem, many decisions were not coded in any category - they were not explicitly prompted nor removed but we also did not believe making them was required to solve the problem. Although most "absent" decisions are not required for solving, some students could decide to put in extra effort and make them regardless. Some decisions that involve more reflection, like D7 (related problems) and D23-26 (reflect), were particularly difficult to identify based on the problem statement unless they were explicitly prompted or removed. Therefore, the amount of prac-
tice with these decisions may be highly student-dependent. Other decisions, such as D1 (importance), D2 (fit), and D27 (implications) are unlikely to be relevant to most undergraduate level problem solving (these decisions were also not universally represented in the analysis of expert problem solving by Price et al. [2]).

IV. DISCUSSION

Our results show that the traditional physics courses (intro, intermediate, and advanced) we analyzed only provide students with practice at making a few key problem-solving decisions - D4 (features), D15 (plan), and D16 (calculations). Other decisions that are important for problem-solving in the real world are either not encountered or are eliminated by being explicitly made for the students. This is mostly due to the nature of the assigned coursework. In an attempt to guide students, curb difficulty, and/or ease grading, instructors may unintentionally remove decisions, reducing the amount of preparation students receive for scientific problem-solving. The courses we analyzed are standard requirements for undergraduate physics majors and span a range of difficulty levels; therefore we believe this work will likely generalize to other universities that have similar courses and assign textbook-style homework problems as the main problem-solving activity. Problems that specify assumptions or have a well-defined expected pathway for solving are particularly likely to remove decisions. However, it is possible to design problems and other learning experiences that involve more decisions, as demonstrated by the two contrasting examples analyzed (capstone and template).

The "template" problems that were specifically designed to prompt problem-solving decisions do indeed give students practice at making many more decisions, though still not the full set. In contrast, a research-focused capstone provided students with opportunities to make nearly every decision from the problem-solving decisions framework. That course was designed as a capstone experience to simulate an authentic research environment; requiring lots of decisions should be a design feature of a good capstone course. Other courses, not analyzed here, may also provide more decision practice, particularly if they involve a research element or include real-world problems that don’t provide all necessary information up front.

Our analysis has some limitations because it is impossible to eliminate the subjectivity of a problem solver during the process of characterizing problem-solving decisions. We attempted to limit inconsistency by discussing a large portion of the problems and comparing student and instructor solutions. We are also in the process of validating our characterization through think-aloud interviews with multiple student solvers. We also stress that our analysis only coded for decisions that were clearly encountered or removed. Theoretically, students could put in a conscious effort to make additional decisions that were absent in our coding. Indeed, the student researcher occasionally did make decisions that we coded as "student-chosen" (not presented in this analysis) because they were not required to solve the problem but the student chose to do extra work (e.g., extra solution checking). Future work will involve think-aloud interviews to analyze which decisions multiple students make when solving these problems in real time.

Since the current curriculum for physics majors typically involves many traditional physics courses and very few research-focused courses (usually as capstone experiences), physics undergraduate students may not be receiving adequate preparation for solving real problems that they will encounter after they graduate. More research is needed to determine an optimal number or sequence of decisions for students to practice in an assignment, but we argue based on the principles of deliberate practice that students will need multiple opportunities to practice most of the decisions. Although it might not be feasible to have an introductory-level student practice all of the decisions in a single problem, they could be supported by gradually increasing the number of decisions required throughout a course. Our work shows that increasing course difficulty does not necessarily increase the number of opportunities students will have to make decisions, so problem-solving decisions need to be explicitly incorporated in the curriculum.

When designing a problem, it is important to consider both how many decisions are encountered and how many decisions are removed. Other research on problem-solving in physics has demonstrated that there can be unintended consequences of eliminating decisions about planning (D15) and representation (D17) [13, 14]. Our analysis of the problem-solving-template problems demonstrates that it is possible to design problems for traditional courses that still provide students with opportunities to deliberately practice a reasonable portion of the problem-solving decisions. In cases where instructors feel students need more support in their problem-solving process, they can opt to prompt a decision rather than eliminate it for the students [15]. Decision prompting has been demonstrated to be a viable approach to incorporate decision-making into introductory physics labs as well [6]. When students are tasked with solving problems in their future careers (whether in physics or not), they will need to make nearly all the problem-solving decisions during their solution process. Therefore, undergraduate programs need to consciously give students more opportunities to practice making these decisions during their coursework.

ACKNOWLEDGMENTS

This work was funded by a Howard Hughes Medical Institute professor’s grant to CEW.


Grappling with the dominant narrative of physics:
Instructors rethink colonial roots together to reshape classrooms

Delwrick Nanthou\(^1\) (he/him)
\(^1\)Department of Physics, South Seattle College, 6000 16th Ave. SW, Seattle, WA, 98106

Ahmed A. Gumale\(^1\) (he/him), Clausell Mathis\(^2\) (he/him), Michelle N. Brown\(^3\) (she/her), and Abigail R. Daane\(^1\) (she/her)
\(^2\)Lyman Briggs, Michigan State University, 426 Auditorium Rd, East Lansing, MI, 48823
\(^3\)Curriculum and Instruction, Pennsylvania State University, 144 Chambers Building, University Park, PA 16802

The colonial roots of science dominate the narrative of physics learning and research. While many acknowledge that the typical physics curriculum does not support students from nondominant groups, it has remained largely unchanged. In an attempt to change this paradigm, a group of instructors voluntarily met to question how one might decolonize a physics classroom. We asked instructors to share whether or not aspects of their identity and pedagogy shifted after participating in the group for two years. In this study, we highlight how identity changed with instructors’ efforts to decolonize their physics curriculum. We also identified instructors describing frustration and exhaustion from systemic constraints in their teaching environment (e.g., student/family pushback, standards, SES). However, within those constraints, each instructor felt empowered to reshape classroom content in an effort to move the community towards more equitable and inclusive physics education.
I. INTRODUCTION

Traditional physics education in the USA has deeply embedded colonial roots that shape what physics content is taught, how we approach teaching and learning, and how we prepare our instructors to teach students. Although efforts have been made to change the cultural practices in physics [1, 2], the education and research community has remained woefully unchanged [3]. In the last twenty years, we see very little improvement in the representation of nondominant groups in physics [4].

All this said, grassroots efforts to move physics instruction towards equity from a small group of secondary and higher education instructors has inspired change in their classrooms. This study examines instructors’ perceptions of their efforts to improve teaching and highlight those changes as an example of productive work in a small learning community. We specifically consider the role of instructor identity in “decolonizing” physics curriculum, which we define as interrogating the Eurocentricity of the physics canon and pedagogy [5]. This study unpacks shifts in instructor identity around equitable instruction through participation in a learning community.

II. ANALYTICAL FRAMEWORK

Instructors’ views on how to act as a physics instructor, what to teach, and how they view their place in society are all associated with their instructor identity [6, 7] and affect what happens in their classroom. Instructor identity has been researched extensively among educational researchers [6]. However, literature specific to the role of instructor identity on equitable physics instruction is scant. Part of the reason for this gap is that both equitable instruction and identity are fluid constructs, making it challenging to connect them explicitly. We define equitable instruction and teaching for equity as teaching in ways that disrupt the prevailing myth of aculturality in physics. This involves establishing meaningful connections between students’ diverse lived experiences and the physics curriculum, with an emphasis on identities and cultures that have traditionally been marginalized or overlooked [8].

To identify and analyze the kinds of shifts the instructors describe, we use C. Mathis’s analytical framework on physics instructor identity towards equitable instruction. We used this framework as a structured lens through which we could discern the kinds of shifts present in the data. The framework includes four instructor conceptions: self, others, knowledge, and pedagogy [8]. Although we recognize how all four conceptions are interconnected and play a role in how instructors conceive and enact equitable physics teaching practices [9], our study specifically considers instructors’ conceptions of self and knowledge.

Conceptions of self describe a physics instructor’s view of themselves and their ability to teach for equity. This includes confidence in their ability to teach effectively, work with administration and colleagues, and support student learning. Conceptions of self can impact perceived instructor competence [10] and instructional quality [11].

Conceptions of knowledge describe physics instructors’ views of canonical (content) knowledge. This includes instructors’ beliefs about the presence of subjectivity in physics. It also includes if and how instructors critique what knowledge “counts” as physics canon (e.g., history and cultural contexts), and their experience with critiquing the content and culture surrounding the physics community.

III. RESEARCH CONTEXT & METHODS

The research context for this study is a professional learning community (PLC) of physics instructors. This voluntary PLC has been meeting online every month for > two years to think about what it would mean to decolonize physics education. They did not previously define what that meant or have a defined plan prior to the start of their work together. The group began with close to 20 instructors attending and over the last year has had eight consistently attending, as well as two additional researchers who observe and participate in conversations.

Each month a different member of the group determines the agenda and leads the 1.5 hrs meeting. The topics discussed ranged from defining and understanding terms (e.g., decolonizing, equity, etc.) to sharing and critiquing lessons, to co-creating new lesson material. Within this context, our research question is: What kinds of shifts come about from this long term, voluntary learning community?

A. Data collection

Our study gathered data from eight instructors across the United States who taught in secondary, two-year college, and university settings. At the end of two years, the instructors were asked several questions about their perceived shifts in their identity and pedagogy. These instructors responded to the initial question, “Can you write or record how your thinking around equitable instruction has changed due to participating in the decolonizing physics group (if at all)”? Five of the eight instructors also participated in a written follow-up reflection asking:

1. What is the goal of the decolonizing physics group?
2. What does it mean to decolonize curricula?
3. How has participation in the decolonizing physics group impacted your views on: yourself, your students, physics content, and your teaching?
4. What will you change going forward after participating in this group?

B. Data analysis

We conducted an open-coding analysis [12] where the first and fifth author sought to identify ways in which the
instructors’ conceptions shifted. The themes were defined and developed based on instructors’ use of descriptive words such as “I realized,” “I learned,” or “I have become.” The first round of coding identified moments where instructors described shifts that aligned with the analytical conceptions. The second author then independently coded the data looking for instructor-identified conception shifts. Next, the three researchers performed an iterative cycle of aligning the statements with the four conceptions. When disagreements appeared, all differences were resolved through discussion. In Section IV, we share themes that were present in at least five instructors’ reflections. We validated the data using extensive member-checking. The eight participating instructors were given the following pseudonyms: Ruby, Layla, Khary, Gwen, Sharon, Bill, Pearl, and Yolanda.

IV. FINDINGS

We found instructors shifted in their conceptions of knowledge and their conceptions of self. While we also found shifts in conceptions of others and pedagogy, these themes were not found in the majority of instructors. The second theme (shared below in IV.B), while not aligned with a particular conception, was added to inform the reader of a more holistic view of the reflections.

A. Instructors’ conceptions of physics knowledge shift

Our first theme highlights a change in instructor’s conceptions of knowledge. All eight instructors share that they changed their classroom content as a result of their work with this group. They each describe content in their instruction that they can or already use to directly challenge the Eurocentric physics narrative. They also found agency in critiquing the historical and contemporary context of their instruction with their students, including different viewpoints of physics into their curricula.

For example, several instructors reflected that they used to view physics from an ideal and theoretical lens and now apply a more grounded, context-rich lens. Bill shares,

“It was eye-opening to think about how the physics concepts I used to teaching, which are so sanitized of any sort of real-life implication, really ought to be presented with the messy, complicated context of our modern world.”

Bill’s imagery of the “sanitized” physics vs. the “messy, complicated context” of physics highlights a movement towards increased real-life connections. Layla compliments Bill’s idea with questions she now uses in lesson planning:

I had never really questioned why we teach certain content beyond the extent of switching up the order, maybe placing more emphasis on momentum (as is done in Matter and Interactions). But I am now realizing how much I had thought as a teacher: “everyone does it this way” or “we’ve always done it this way” and I’m questioning it... I am now in the habit of asking myself as I plan each unit: why is this science important, how is it used today, how can I encourage my students to find connections.

Bill and Layla both describe a shift from teaching on auto-pilot and how they have started to think deeper about why they are teaching, what they're teaching, and how.

Getting a bit more specific to the classroom, Ruby shares some examples of how that has played out:

I used to try to downplay history because I didn’t want to remind people that all of the famous physicists are white men, but now, I talk about that being problematic. I explicitly mention that one man got credit for many people’s work - often at the expense of others' lives. For example, I discuss how Newton was heavily invested in the slave trade or that Ibn Sahl came before Snell and Snell still gets the credit.

Ruby acknowledges that it is important to represent the power dynamics involved in recognizing contributions of certain individuals to science. She also talks about people who have acquired fame for their work in science but were also involved with activities that were harmful to others.

Still other instructors discuss how they have learned about more historical context about certain topics. For example, both Sharon and Pearl use batteries as an example of content they have shifted. Sharon states,

“I’ve learned new content (like the Baghdad battery or the more modern history of colonial oppression in batteries), and I’ve been pushed to think deeper about overarching frameworks.”

Pearl shares that she now teaches these ideas explicitly,

“Thanks to this group, I have tried incorporating pieces of lessons like the conflict mineral in the electric battery industry, and the structure of how I elicited student ideas about inertia.”

Pearl highlights shifts in her own teaching that directly affects the students in both content (battery-focused ethical issues) and practice (eliciting student ideas).

Yolanda also mentions batteries as an example of content change, but also shares the kinds of tools that the group has given her during their discussions.

I had always been uncomfortable teaching about projectiles, nuclear reactions/weapons, and many other topics that so glaringly had to do with war and killing but I didn’t know what to do with that. This group has offered me tools, frameworks, protocols, language, and confidence to speak openly with my students about the ways in which physics colonizes and to offer ways in which physics can be used for GOOD. Some of my favorite lessons have come from work done in this group (ie: batteries, climate change, renewable energy, recognition vs contribution in physics).

Not only does Yolanda describe several examples of content change, she also expands on how the current context in which physics is used can be extremely problematic. At the same time, she has found comfort in
learning about how to articulate the “ways in which physics can be used for GOOD.”

B. Instructors feel bound by the academic system

Instructors at all levels share constraints across multiple contexts that do not support them in decolonizing physics in their educational environment. While they share this theme, we do not see instructors in this group enacting prior research highlighting “race evasiveness” [13] or “system-blaming powerlessness” [14]. During their work, they reflect that there are both individual and systemic issues outside of their sphere of influence that affect their ability to change their courses: the classroom setting, physical location, standards, colleagues, and student families. For example, Bill writes about the kind of classroom as restricting his teaching:

I have benefited a lot from the discussions in this group, even as I have struggled to bring ideas back to my lecture-based, large-enrollment college physics classes and labs...I’ve become quite a bit pessimistic about the possibility of meaningful physics instruction in large colleges...I am mostly treading water and trying to help the students get through the system as well as they can.

Here, Bill describes the lecture-based teaching approaches as restricting the kinds of changes that can be made. Their description of getting students “through the system” and “treading water” seems to imply that some aspects of the power lies in the system/water and not with the instructor.

This idea of a lack of empowerment also presented itself in terms of depending on the specific state standards of the instructor as Gwen described here:

I did not realize how bound the teachers in this group feel...With teachers in [my state], I am frustrated by the standard, ‘the standard says...’ comes up. (Evidently [certain USA states omitted for anonymity] are the most bound to standards.) Do we need to change the standards? How do we do that?

Layla shares a similar idea that it is the type of course (and presumably the corresponding standards) that constrain her:

“Unfortunately, I teach AP courses so I can’t switch things up as much as I’d like.” Khary mentions that even within her school, individuals affect her agency in her teaching:

It’s been somewhat challenging to do this work well for me; there’s a bit of idealism in it which I experience as the anxiety of waiting for a student/family to push back on why this is happening in my class, particularly when no other colleagues are doing the same thing. I don’t want to be on the frontlines any longer because it’s psychically draining for me.

While the majority of the instructors shared more systemic academic traits, Khary’s reflection highlights that even within the school, the voices of students, their families, and colleagues can affect instructors’ sense of agency to significantly change their practice. At first glance, this might appear as a more individualized and anecdotal example, but this kind of pushback from students has been identified in other contexts [15].

C. Instructors’ conceptions of knowledge shift regarding the idea of “decolonizing physics”

Each instructor explicitly writes that they cannot yet, and do not yet know how to, decolonize physics. While the group may have started out trying to decolonize, all these instructors articulate that the efforts of the group have shifted to smaller scale endeavors: they are not trying to completely overhaul the curriculum yet, but rather make it more equitable. Seven of the eight instructors in the group described a recognition that “decolonizing physics,” as the group had set out to do, was not yet achievable for themselves. They acknowledged that they needed to do more research and learn more about the topic so they could effectively teach it to their students. For example, Ruby wrote, “[Decolonizing curricula] is maybe not possible to do!” Bill wrote that the group's efforts are not focused on decolonizing physics, but rather for truth and equity within the curriculum.

I am not sure that this [decolonizing physics] is something that the group is doing. Decolonize is a heavy word, and there’s a high bar. We are maybe trying to improve our instruction to be more truthful and equitable."

In a follow up conversation during our member-checking process, Bill elaborated that the group was not decolonizing in the sense of “substantial reparations for the descendants of chattel slaves” or “land transferred back to First Nations”. This is another example of a shift in conceptions of knowledge in terms of equitable instruction.

D. Instructors’ conceptions of self shift: they realize how much they can still learn and stay motivated

Instructors’ conceptions of self shifted towards an awareness of limitations, while also increasing their willingness to continue expanding their knowledge. In their reflections, every single instructor described how they are still in the process of learning about how to better teach physics. Khary shares, “I’ve learned how little I actually knew about how knowledge is constructed within a STEM discipline, and how much I took for granted that everything I learned was objective.” Just like Khary shared in the previous quote, other instructors identified that the scope of decolonizing physics was much larger than they initially anticipated it to be. For example, Pearl states that she is “still emerging” in her skills about “knowing how to facilitate class conversations digging into this kind of content.” Layla also shares something similar and goes on to describe how that shows up as a new way of interacting with her students:

While I can’t claim to be very knowledgeable yet, I at least am at the point in the Dunning-Kruger effect where I know that I don’t know and I feel comfortable with
That. I am happy to share with my students “Look, I don’t know this. Let’s explore together.” That’s been a big shift for me.

Layla describes her learning using the Dunning-Kruger effect, where experts in a particular field are more likely to identify gaps in their knowledge than novices who may not yet recognize those gaps. Layla shares that she does not yet feel “very knowledgeable,” but she is allowing herself to be more vulnerable with her students. She now models that it is okay to not know everything, exemplifying a learning mentality to effectively better serve her students. Gwen’s reflection showcases an understanding that she must put in deliberate effort to first unlearn problematic ideas to then move towards a more equitable framing of physics.

While in this group, I realized that the implicit message of race and whiteness associated with authority in physics had also been ingrained in me, and that, like anything, it will take explicit unveiling, reflection, interrogation, to uncover and disturb the assumption that physics belongs somehow to white people and especially white men.

Gwen articulates that while she engaged with the group, she discovered that she had unknowingly internalized the eurocentric narrative of physics. Understanding that this message is “ingrained” in her, she expressed a need for deliberate efforts to “uncover and disturb the assumptions.”

Even while they all recognize that their work is unfinished after two years, instructors also write that they still value the meetings and are not deterred from the work - even if they may never fully create a fully “decolonized curriculum”, as shared in Section III.B. It is also noteworthy that all of the instructors came to this place even though they came into this project with rather different views on what it would mean for physics instruction to provide more equitable outcomes. For example, Ruby shares, “I think of this as more of an ongoing, lifelong endeavor...There is still so much to learn - so perhaps it is not a shifted view so much as increasing the depth and commitment to those views.” Other instructors write explicitly about how the group is a reason for this change. For example, Gwen shares, “I’m happy that I’m still receiving big a-ha’s during [these meetings].” Pearl expands on the feeling of ongoing effort, even as she reflects a sense that she feels she is just beginning the work, I can tell I’m just at the start of decolonizing the bulk of my physics curriculum, but thanks to this group I feel like I can see a fuzzy road map for how to do it. Our conversations make me feel empowered and also give me a sense of urgency towards making changes happen in my own classroom.

Pearl writes that though she is “just at the start,” she sees a “fuzzy road map” for how to decolonize her curriculum, feeling “empowered” to make “changes happen” in her classroom. Along these lines, Yolanda also writes of empowerment: “This group gives me hope and every meeting leaves me feeling energized and empowered.”

V. DISCUSSION & CONCLUSION

Our analysis of instructors’ written reflections identified four major themes. First, this group co-revised and expanded their content in a way that made them see the curriculum as more equitable, thus refining their conceptions of physics knowledge - and how that applies pedagogically. The content shift (e.g., climate change, battery origins, Newton’s problematic actions) are a concrete, albeit relatively safe, example of how instructors have changed aspects of the curriculum, without having to deal with the negative consequences of school community pushback and the overarching structure of colonization.

Instructors also described feeling bound by their educational systems as they attempt to decolonize physics. The system, in this case, does not change structurally (e.g., the lecture hall, the inequitable representation, private schools vs. public, the inequity outside the classroom itself). Through their work in this group, these instructors developed an understanding that they cannot change the system as a whole (yet!), but they can move the needle towards equity. We saw their reflections as demonstrating several aspects of transformative professional learning [16] and see this study as an initial step towards articulating one example of how physics instructors might productively engage in moving towards more equitable instruction.

Our fourth theme speaks to instructors’ feelings of productivity and hope. We are encouraged by their desire to continue their efforts. Undaunted, instructors found ways around these circumstances through gradually making changes to curriculum as shared in Section IV.A. Through the conceptions of self framework, we see that instructors have strengthened their sense of empowerment to do this work. Even though these instructors came in with different viewpoints, their interactions and support for each other’s learning has created more motivation and action.

A major implication of this study is that this PLC shifted the canon by revisiting, problematizing, and/or telling a truer story of history. For equity work in physics education research, this study highlights an example of what physics instructors can do given the opportunity. We recommend looking to instructors for expertise in developing more equitable curricula and partnering with them to identify additional tools to connect with students and share different perspectives. This partnership will ultimately create stronger curricula and healthier, more diverse, and more inclusive classrooms.

ACKNOWLEDGMENTS

We are grateful to the PLC group for their generosity in making their meetings and ideas accessible to the team. We acknowledge the Culture-Based Phys. Educ. Res. Group at Michigan State University for their support.
Supporting first- and second-order departmental change with the Effective Practices for Physics Programs (EP3) Guide

Christine O’Donnell (she/her)
American Physical Society, One Physics Ellipse, College Park, MD, 20740

Stephanie Viola Chasteen (she/her)
Chasteen Educational Consulting, Louisville CO, 80027

Many reports, research, and initiatives have presented evidence-based strategies to create strong departments. The Effective Practices for Physics Programs (EP3) Initiative is a collaborative effort between the American Physical Society (APS) and the American Association of Physics Teachers (AAPT) to compile and curate such resources in an online “Guide” for departments to empower high-quality physics education. To ensure the Guide helps improve physics education, this work aims to understand whether the Guide (a written document) can effectively support departmental change efforts (which can be complex). We draw from findings from a 2020 survey (N=310), 2022 survey (N=239), and 2022 interviews (N=8) of physics department chairs at colleges and universities in the US. In the survey, 22% of respondents had used the Guide and 18% had plans to use it, e.g., for strategic planning. Our interviewees spoke about their limited ability to engage in anything that is not immediately urgent and/or requires a significant investment of time. However, many also talked about getting good ideas from the EP3 Guide and using the EP3 Guide in a strategic way, such as during faculty meetings to spark discussion. We find that among EP3 Guide users, the Guide is a potentially effective tool to support first-order change, i.e., change that works within existing systems and worldviews, since interviewees reported using content from the EP3 Guide to understand, frame, and promote their departmental change efforts. These successes can also be leveraged to potentially engage the EP3 audience in second-order change, i.e., change that requires reframing goals and/or values, developing new structures, or other transformational processes. However, additional active supports such as webinars, short courses, or leadership institutes may be necessary for effective and sustained second-order change.
I. INTRODUCTION

A long-standing challenge faced by academic departments in higher education is how to create and sustain thriving programs that effectively educate diverse students. For physics departments, these challenges can have added urgency due to resource limitations (e.g., those generated by the coronavirus pandemic) or existential threats of being closed or merged with another department [1, 2].

Identifying strategies to help physics departments thrive is not a new effort. The Strategic Programs for Innovations in Undergraduate Physics (SPIN-UP) report [3] emphasized that strength arises from within: thriving programs take responsibility for their own growth, continuously evaluate their progress, and provide a strong education for their students. While the SPIN-UP report was transformative for many in the physics community and thus is an important milestone, one major drawback is its static nature. In the 20 years since its publication, the landscape of physics education has continued to change. Departmental growth has slowed, new teaching needs such as computational physics have arisen, and departments are under increasing pressure to more effectively and comprehensively address equity, diversity, and inclusion (EDI) [4].

To meet these needs a new initiative has been spearheaded: the Effective Practices for Physics Programs (EP3) Initiative. EP3 is a collaborative effort between the American Physical Society (APS) and the American Association of Physics Teachers (AAPT). The foundational product is the EP3 Guide (ep3guide.org), which is a comprehensive living document that compiles evidence-based strategies. The Guide currently includes 25 sections on topics including recruitment, retention, EDI, computational skills, and more. Each section was written and reviewed by members of the physics community including physics faculty (including PER), university administrators, and industry professionals; to date, over 240 people from more than 140 institutions have been a part of the Guide development process. Guide content will be regularly revised to ensure it remains current, with new sections added as appropriate. The Guide also emphasizes using a cycle of reflection and action, which encourages users to be intentional as they navigate the nonlinear nature of departmental change.

In addition to the written Guide, the EP3 Initiative has also developed programming designed to actively engage departments in change efforts. For example, EP3 has created the Departmental Action Leadership Institute (DALI), which has led 4 cohorts of physics faculty from 19 departments to engage with change leadership skills and practices. Participants reported that DALI increased their motivation and confidence in enacting change efforts and led to more measured approaches that included broad stakeholder involvement [5].

However, DALI can only serve a small portion of all physics departments at a given time. A written Guide has the ability to reach a larger audience to support additional change efforts, and so we seek to understand whether and how a written Guide – regardless of whether it is a static publication or updated online content – can lead to deep change, and if so, whether some kinds of changes are better supported by the written document. After all, students don’t learn physics from a textbook without a lot of support, so how can departments learn effective change strategies just from written advice?

The analysis presented here draws on findings from EP3’s external evaluation. In §II, we briefly review departmental change literature to provide an interpretive framework. We describe the evaluation activities in §III and present the findings in §IV. In §V, we discuss implications and recommendations for EP3 and other departmental change efforts.

II. LITERATURE REVIEW

Scholars have proposed many different frameworks to describe change efforts, including within the context of higher education. One common classification scheme is to differentiate between first-order and second-order change based on what is being affected by the change effort. First-order change refers to change that works within existing systems and worldviews. Second-order change can be seen as transformational or radical change because it involves a fundamental shift in values, goals, worldviews, missions, and/or structures [6]. For example, imagine a physics department determines that students in the introductory physics courses are struggling when taking calculus at the same time. A first-order change might be to connect students with on-campus tutoring options so that they can receive additional instruction outside of class. Second-order changes, on the other hand, would include reforming introductory courses so that they provide opportunities for students to learn and practice math skills.

Researchers have applied the framework of first- and second-order change to STEM and physics change efforts. For example, [7] reflected on efforts to implement active learning practices, noting that while many instructors are aware of these practices, many instructors do not use or discontinue use of active learning [8, 9]. They critiqued efforts that largely relied on the inherent value of active learning rather than addressing “deep-rooted institutional structures and cultural norms” that discouraged its use; they argue second-order change is critical. Similarly, [10] critiqued change efforts that focus on individual faculty practices as “overly simplistic” and unable to create widespread sustained change. Separately, [11] summarized the TEAM-UP report’s [12] argument that cultural change is necessary to address why the number of physics bachelor’s degrees has more than doubled for all federally reported racial and ethnic groups except Black or African Americans (as well as American Indian and Alaskan Natives). Because the physics community needs to grapple with its own culture, value, and norms, using only first-order methods is likely to end in failure [11].

In this paper, we use this framework of first- and second-order change in our examination of data on EP3 Guide use to explore whether and how a written guide can support change.
III. BACKGROUND AND METHODS

This paper draws from data collected for the external evaluation of the EP3 Initiative in 2020-2023. We focus on findings from interviews with department chairs, which builds from the results of surveys of physics department chairs.

A. 2020 and 2022 surveys of physics department chairs

In May 2020 and again in June 2022, a survey was distributed to all physics department chairs in the US. The 2020 survey focused on departmental challenges and culture [1] and received 310 responses (41% response rate; 30% of respondents were from Ph.D.-granting, 9% M.S.-granting, and 61% B.S.-granting institutions). The 2022 survey focused on use of the Guide [2] and received 239 responses (32% response rate; 25% of respondents were from Ph.D.-granting institutions, 8% M.S.-granting, and 66% B.S.-granting institutions). These fractions were generally representative of the institutions invited to participate. In the 2022 survey, the top problem reported by departments was low enrollment in the major (71% of respondents), but this was less commonly a challenge for Ph.D.-granting institutions (47% cited low enrollment as a challenge). Other problems cited in both surveys were inadequate preparation of incoming students and low enrollment or retention of historically marginalized groups. Additionally, in 2022, the majority (76% of respondents) reported facing a moderate threat in the last 2 years, such as low enrollments (55%) or a reduction in faculty lines (32%); these rates were higher than the 2020 survey [1]. Further, 19% reported facing a severe threat, such as a department merger or closure. B.S.-granting departments were more likely to report severe threats. Additionally, most (72%) of respondents said they were aware of the EP3 Guide, and 31% of these reported that they used it in some way (22% of all survey respondents). For those who have used or plan to use the Guide, nearly half describe strategic uses, such as to guide strategic planning or program review [2].

B. Interviews of physics department chairs

To learn more about Guide use, 2022 survey respondents were invited to participate in follow-up interviews. Out of the 39 who volunteered to participate, 11 were excluded because they had not heard of the Guide or had served as Guide contributors. From the remaining 28, nineteen potential participants were selected, prioritizing those who had actually used the Guide. Eleven declined or did not respond to the request; thus a total of 8 department chairs were interviewed over Zoom between December 2022 - February 2023. Because few Ph.D.-granting institutions reported using the Guide, and several of those who did declined the interview request, these criteria resulted in an undersampling of Ph.D.-granting institutions. Of these 8 chairs, 5 had used the Guide and 3 had concrete plans to do so; 7 had experienced moderate threats, and only 1 reported a severe threat. Five were at B.S.-granting institutions, two were at M.S.-granting institutions, and one was at a Ph.D.-granting institution. All were at institutions of ∼10,000 or fewer students, and 5 were at institutions with fewer than 5,000 students. We did not collect personal demographic data, but the majority of participants presented as white men.

Interviews followed a semi-formal protocol covering use and perception of the Guide, experience of departmental threats, and advice for the EP3 Initiative. Guide users were asked to describe their latest or most extensive experience with the Guide in detail, including where they began, who was involved, and how ideas were captured. Interviews were conducted by Author SVC, with Author CO attending most interviews; both participated in interpretation of findings. While a fuller discussion is beyond this paper’s scope, we recognize that our positionalities inherently influence all interactions we have, including these interviews. Author CO is a Program Manager on the EP3 Initiative with a background in PER/DBER and presents as a White-passing cis-woman. Author SVC serves as the External Evaluator of the EP3 project and identifies as a physicist, consultant, and white woman.

IV. INTERVIEW FINDINGS

Here we present major findings from our interviews. In the following section (§V), we analyze these findings and connect them to the framework of first- and second-order change.

A. Interviewees experienced situational overwhelm

Many (N=3) of the interviewees, especially those at small institutions, were completely overwhelmed. Additionally, several declined the interview request due to time constraints despite a desire to help the project. We heard from chairs who felt shame over having to focus only on immediately urgent issues rather than also being able to address important but non-urgent issues in their department. Some of these chairs were at universities with fewer faculty lines yet having to teach just as many courses, leading to overwork for themselves and their colleagues. Any change efforts will need to account for the situational overwhelm that some chairs face.

B. Regular reminders help with busy professional lives

Three interviewees, whether or not they were overwhelmed, indicated that they benefit from regular reminders. There are many initiatives in physics and education, and reminders were welcomed to keep resources on the top of their minds. Four interviewees expressed a need for small, manageable, and targeted opportunities to engage with EP3 ideas without inundating them. Thus, email reminders, newsletters, and webinars, along with specific examples of how others have used resources, may increase Guide uptake.
C. The Guide content and headings serve as a gateway to understanding a topic

A few interviewees indicated that they perused the Guide content, or heading structure, to learn about the topic (e.g., self study). For example, one interviewee talked about using the Guide structure to understand the taxonomy or dimensions of recruiting students. The Guide “[broke] it down into some sort of big categories for you.” They also used the Guide to understand the specifics of these practices, mentioning that “for me, one fear is that we’re going to do something and forget some key element of it,” and the Guide helped to alleviate this concern. This orientation was valuable for interviewees, but they did not capture ideas systematically (e.g., in notes) and did not seek a way to do so; instead, they expressed that the Guide content went into their general pool of knowledge. This pool of knowledge enabled them (in some cases) to plan their future action or engagement with other faculty. We recognize, however, that the lack of systematic capture of ideas reduces the potential for circling back to possible action items, reflecting on how efforts went, or gaining new (second-order) perspectives on the program.

D. There is a tendency to use the content “just in time”

Several interviewees reported that they used the Guide because of specific initiatives:

- 2 interviewees used the Guide for curriculum reviews,
- 1 interviewee used the Guide while developing a new degree track in engineering physics and planned to use the Guide when reviewing upper-level courses,
- 1 interviewee was a new chair and read the section on how to be an effective department chair,
- 1 used the Guide to serve as an external reviewer for another physics department, and
- 1 used the Guide to prepare for a College-wide presentation about EDI to fellow science faculty.

These situations can be leveraged as starting points for engaging Guide users. These common needs for “just in time” resources also underline the value of regular reminders of resources like the Guide, so that a chair is reminded of key resources at opportune moments such as the start of a new grant. At the same time, “just in time” use is not necessarily geared to addressing strategic priorities or program gaps.

E. Many focused on familiar ideas and current practice

Many interviewees described Guide use that seemed to focus on identifying ideas that aligned with what they were already doing or planning to do. For example, one interviewee focused on Guide headings that related to a recruitment strategy they were already using. Another interviewee read the Guide to confirm whether they were on the “right track” and should continue doing certain efforts. Additionally, because the content in the Guide can be overwhelming, starting with familiar content was a way to reduce that overwhelm. Lastly, for a busy chair, a focus on improving the things they are already doing could be more manageable than trying to start something completely new. Unfortunately, that means many of our interviewees were not using the Guide as a way to find new strategies and become exposed to new ideas. This behavior also limits the ability of the Guide to lead to second order change.

F. The Guide was sometimes used in strategic planning for the department, driving faculty conversations

In both our survey and interviews, respondents used the EP3 Guide to support long-term strategy and planning. Two interviewees plan to use the Guide to support their future program reviews. Additionally, one interviewee described plans to use the Guide to implement priorities from a new departmental strategic plan. We also heard from several interviewees who used (N=4) or plan to use (N=1) the Guide to support faculty discussions. One person said the Guide was an effective “jumping off point”, with another interviewee describing the discussion as a generative exercise that used the headings in the Guide section about recruitment as “major bullet points”. These methods allowed them to use the Guide as a tool for generating action items for the department. Furthermore, two interviewees specifically mentioned the enhanced clout of resources promoted by organizations like APS and AAPT, and one interviewee specifically shared the the Guide with their faculty as “[the Guide is] supported by APS as the ‘right way’ to build an effective physics program.”

These findings mirror findings from the 2020 survey [1] that departments do engage in productive strategic planning. In the 2020 survey, between 63% and 70% of respondents agreed that changes to the undergraduate program were driven by departmental goals (as opposed to outside threats); were driven by the department’s purpose; were seen as ongoing processes; were driven by shared responsibility among the faculty; and were supported by an innovative, experimental, and learning-oriented departmental culture. Additionally, the majority agreed that program review is approached as an opportunity for improvement (90%) and resulted in positive change (79%). However, there was room for improvement in such strategic change processes: only 37% agreed that program changes involve multiple stakeholders, and only 42% agreed that program changes are supported by data and/or assessment results, even though they would like them to be. Lastly, only 65% indicated that external review results were revisited periodically to guide the department [1].

In one interview, we discussed how we can better support their use of the Guide for strategic planning, e.g., by providing reflective prompts for faculty as they read the Guide. These findings can inform future resources created by the EP3 Initiative. Additionally, a request that came up in both the survey and in two interviews is for the Initiative to provide case studies about Guide uses. The EP3 Initiative should make sure Guide case studies include these strategic uses.
Although many of our interviewees were facing situational overwhelm, they generally wanted regular reminders about the resources available through EP3. Those who had used the Guide (as well as the larger set of survey respondents) found it helped them to build out their understanding of a topic’s dimensions, and they tended to use guide content “just in time” and for strategic planning. We also note the particular positioning of the EP3 Guide – a document backed by APS and AAPT that was developed with broad involvement of the physics community – gives enhanced clout to its recommendations. In the 2022 survey, most respondents “moderately” or “strongly agreed” that the Guide is relevant (62%) and valuable (61%) to their departments; they also generally reported that “some” or “most” of their faculty would agree as well (72% for relevance and 73% for value) [2].

Thus, we find that the written guides like the EP3 Guide are potentially effective tools to support first-order change, which can be more familiar to change leaders as an accessible first step towards navigating overwhelming and complex change processes. Similarly, the 2003 SPIN-UP report [3] was taken up by the community as a “call to arms” to address the challenges and opportunities faced by physics departments [4]. Our findings suggest the EP3 Guide can be the next iteration.

We also note parallels between these Guide uses and research-based principles about learning. For example, how learners organize information influences their ability to apply that knowledge [13]. The EP3 Initiative was very deliberate in the Guide’s structure to provide a clear organization. Interviewees talked about using the Guide to understand the dimensions of an issue (§IV C), and this kind of learning can be a valuable start for larger change efforts. Learning is also a social activity [13], and we observed a tendency to use the Guide to spark discussions (§IV F). However, [13] also emphasized that learners need to monitor and adjust their approaches to learning to become self-directed learners. None of our interviewees systematically captured the ideas generated from the EP3 Guide (§IV C), hindering their ability to return to these ideas. Thus, a limitation of a written Guide is its inability to support metacognition and reflection, both of which are necessary (though not sufficient) to create effective change efforts, particularly second-order change.

Additionally, these parallels also suggest potential obstacles for deeper, second-order change from Guide use. Learners’ prior knowledge significantly affects learning processes [13], and we observed that interviewees often started with content that aligned with current knowledge and initiatives. However, if a user filters Guide content based on their own opinions and worldviews about what is important or productive, they could unintentionally limit their ability to engage with second-order change. As a parallel example, on the 2020 survey [1], the vast majority of survey respondents (74%) indicated that low enrollment or retention of historically marginalized individuals was a top problem for their department. However, only about half of these respondents indicated that department climate and creating an inclusive learning environment was an issue, suggesting a gap between perceived problems versus the change efforts needed to address systemic and cultural barriers. A written guide has limited ability to address these gaps; the deeper learning and broader stakeholder engagement needed for second-order change cannot be addressed with only first-order methods [11].

That said, written documents like the EP3 Guide can be a gateway to engaging users in second-order change. Survey respondents and interviewees appreciated the EP3 Guide’s carefully organized structure, research- and community-based advice, and coverage of challenges that real departments have. Using content from a written guide can serve as an introduction to change efforts, growing individuals’ sense of competence and agency. For the EP3 Guide in particular, its backing from both APS and AAPT provides additional clout to related change efforts. Interviewees talked about using the Guide to confirm they were on the “right track”; we recommend that the EP3 Initiative develop case studies to recognize and celebrate Guide-aligned first-order change efforts.

Finally, we recommend that the EP3 Initiative should continue to leverage these uses to design and assess additional activities that support second-order change. For example, EP3-facilitated leadership institutes encourage users to adopt a cycle of reflection and action that involves broad stakeholders and cultural change [5, 14], and EP3 should explore whether workshops, webinars, or other formats can create similar outcomes. However, because we recognize some chairs and change leaders face significant situational overwhelm, the EP3 Initiative should be mindful of participants’ contexts when creating these activities. Further, these activities can be designed for departments at different stages of change efforts. For example, in the 2020 survey, 70% of respondents said they had a strategic plan [1], meaning workshops, etc. about developing a strategic plan could target the roughly 30% of departments without one. Furthermore, only 41% of those with a strategic plan said they felt that they used it productively, so another set of workshops, etc. could be focused on how to effectively use and evaluate strategic plan progress. By creating pathways that empower change leaders to begin with first-order change inspired by a written document and moves them towards more involved engagements that support second-order change, the EP3 Initiative and other departmental change efforts can effectively support departments to sustainably improve and provide high-quality physics education.

ACKNOWLEDGMENTS

We thank Chandra Turpen for helpful feedback and conversation. This work is supported by NSF under Grant Nos. 1738311, 1747563, 1821372, 2033894 and Chasteen Educational Consulting. Author SVC is the external evaluator for EP3 and conducted data collection & analysis in that role, but paper authoring was not part of the consulting agreement. Finally, we thank the anonymous reviewers for their feedback.
[14] D. Sachmazidi et al. (in prep.).
Three axes for expressing disability models and experiences: The cause, the effect, and the ability/disability dichotomy

Daniel Oleynik (He/They)
Department of Physics, University of Central Florida, 4111 Libra Drive, Orlando, FL, 32816

Constance M. Doty (She/Her)
Department of Physics, University of Central Florida, 4111 Libra Dr., Orlando, FL, 32816

Erin M. Scanlon (She/Her)
Department of Physics, University of Connecticut – Avery Point, 1084 Shennecossett Rd., Groton, CT, 06340

Jacquelyn J. Chini (She/Her)
Department of Physics, University of Central Florida, 4111 Libra Dr., Orlando, FL, 32816

In interviews with physics students and early career physicists, we ask about their experiences with having impairments in the physics setting and physics culture. In this paper, we highlight how experiences shared by participants as disabled people in physics represent clusters of models of disability. Specifically, we apply a theoretical framing of a three-dimensional disability model space, with axes defined as medical versus social (i.e., cause); tragedy versus affirmative (i.e., effect); and minority group versus universal (i.e., ability/disability dichotomy). For example, in this framework, providing accommodations is described by a cluster of the social and minority models of disability. By analyzing participants’ experiences in physics through this disability framework, we aim to identify the models that underpin supportive experiences and support the development of policies and professional development for the physics community towards benefiting disabled people. Through analysis and comparison of these models and participants’ narratives, we offer a discussion and possible guidelines for instructors interacting with students with disabilities, opportunities for those with disabilities to deconstruct their own prior experiences and analyze potential misinterpretations that may arise from the models.
I. INTRODUCTION

In postsecondary education, disabled students represent approximately 20% of the student population [1,2]. With such a sizeable population, it is important to design curriculum and courses with disabled students in mind. Yet, many physicists are unaware of the life experiences of disabled students [3], which may make it difficult to empathize with current students within their courses. By highlighting students’ lived experiences through different models of disability, we hope to increase instructors’ capacity to relate to their students, to support instructors to become better mentors, and to implement accessible and inclusive practices in their teaching and research interactions. The culture of the physics community is inherently ablest due to its valuing of ability; overall “academia powerfully mandates able-bodiedness and able-mindedness, as well as other forms of social and communicative hyper-ability” [4] (pg. 7). Interactions in higher education are typically shaped by the medical model of disability, under which disability is construed as biological and individual in nature. The goal of this paper is to present a range of models that can be used to analyze interactions around disability and to use these models to describe disabled students’ experiences with physics instructors.

In this study, we use six disability models outlined by Goldiner [5]. Goldiner highlights that the medical and social models vary in where they situate the cause of disability. The tragedy and affirmative models vary in the effect of disability. Finally, the minority model defines a threshold between those who are disabled and those who are not, while the universal model removes the threshold and posits that everyone exists on a spectrum of disability. We used these models to analyze descriptions of interactions between disabled physics students and physics instructors during interviews conducted with disabled physics students.

II. POSITIONALITY AND LANGUAGE

The identities of the authors change how we conduct our research and interact with the interviewees. The members of the research group experience a variety of impairments, including emotional/mental health, physical, hearing, and health impairments. Use of person-first (e.g., person with disability) and identity-first (e.g., disabled person) language varies depending on the context and the person within the disability community [5,6]. The first author’s preferred identity label is Dis.,1 due to its separation of the concept of ability from the author’s identity. However, participants also have their own preferences regarding language about their identity. Thus, we will use the interviewee’s preferred language when discussing their interviews but will use “Dis.” elsewhere. We also intentionally distinguish between disability and impairment. Disability will be defined based on the corresponding model, while impairment will describe biological difference from the norm [6].

III. MODELS OF DISABILITY FRAMEWORK

Goldiner positions six models of disability to construct a three-dimensional space, which allows an interaction to be described by a cluster of models. Below, we provide more details about and examples for each of the six models [5].

A. Medical versus Social

The medical and social models revolve around a debate about the cause of disability. Goldiner views impairments as the “cause of their disadvantage in social participation” [6]. The medical model suggests that disability is something to be cured or fixed through engagement with the medical community (e.g., medications or assistive technologies); the burden for change is placed on the Dis. person. An example of the medical model in action are state-sponsored disability benefits where disability is defined as “inability to engage in any substantial gainful activity by reason of any medically determinable physical or mental impairment” [5]. In the medical model, a Dis individual is deemed not able to work because of their impairment, without attention to the interaction between a person with impairments and potential environments.

The social model defines disability as caused by societal and physical barriers to access. Goldiner cites The Union of the Physically Impaired Against Segregation, stating “[I]t is society which disables physically impaired people. Disability is something imposed on top of our impairments, by the way we are unnecessarily isolated and excluded from full participation in society.” [5, 7] The burden of change in the social model is the physical and social environment. For example, if roads and buildings were designed while accounting for those who use wheelchairs (e.g., curb cuts), those individuals would have the opportunity to operate with less disabling barriers. [5] For example, during the pandemic, those who were unable to leave their house due to their impairment gained more access to work as society shifted towards providing virtual options. [8]

B. Tragedy versus Affirmative

The tragedy and affirmative models are based on the effects of impairments. The tragedy model rests on the claim that disability and impairments generate tragedy and harm. At the extreme end, the tragedy model posits that a disabled

1 1 – Dis. is the author’s personal disability identity because of its separation from the idea of ability and identity, and an attempt to reclaim the use of “dis” being negative
life is a life not worth living. For example, the belief by some in the deaf community that being ‘deaf is not disabled’ [5] is predicated within the tragedy model as some view deafness as a human variation with positive value rather than a loss caused by impairments [9, 10]. This framing positions “positive value” outside of disability.

On the opposite side of the spectrum exists the affirmative model. The affirmative model highlights positive aspects that impairments may allow. At its most general, the affirmative model describes the potential for the effects of impairments to be beneficial. One example of this is how neurodivergent individuals often also identify as queer, possibly because their disability allows them to reject social conventions and gender norms [5, 11].

C. Minority and Universal

The minority model espouses the belief that there is some separation between those who are disabled and those who are not disabled. The threshold for what allows for one to identify as disabled varies on the context and one’s own personal beliefs. The existence of what constitutes the threshold varies, ranging from a physical marker of disability, a feeling of weakness, or a sense of belonging in a community. Dis. individuals may have feelings that they may not belong in the community due to their own belief about where the threshold between disabled and not disabled lies. Additionally, similar circumstances may fall under separate sides of the threshold. Bagenstos [12] argues that between two individuals who use wheelchairs, one with paraplegia and one with a broken leg, the individual with paraplegia is more likely to be considered disabled.

Conversely, the universal model is predicated on the belief that anyone can be disabled, and that everyone falls somewhere on the disability spectrum. Rather than a subjective threshold, the universal model does not delineate between disabled or non-disabled, but rather that disability is part of the human condition. Goldiner states that some theorists view disability “on a lifetime spectrum, rather than at a specific point in time. On this view, if almost everyone will experience disability at some point in their lives, then the fact that the risk has not been realized for everyone at a particular point in time does not undermine the universality of disability” [5].

III. METHODS

We recruited volunteers for pilot interviews through personal contacts with the authors (convenience sample). Participants shared demographic information which included their gender identity, race, and personal disabilities that they identified with via an online survey before the interview and indicated their choice of interview format (i.e., a single one-hour interview, two one-hour interviews, or one two-hour interview). Participants were invited to share their access needs for the interviews, such as requesting a sign language interpreter or an asynchronous interview format. All interviews were conducted virtually by the first author via Zoom. Participants were provided a gift card valued at $25 for a one-hour interview and $50 for two-hours of interview.

Three semi-structured interviews were conducted with Dis. students about their experiences in physics learning and research environments. During the interview, we asked participants about the disabilities they identified with, their experiences of those disabilities, and the relationship between disability and their identity. Afterwards, we asked about the supports and barriers they experienced within the physics community in both academic and research settings. Finally, we discussed what an inclusive physics community and classroom would look like, in their opinion, with suggestions for prospective instructors and advisors to make an inclusive and accessible community for Dis. students.

When analyzing the verbatim transcripts of the interview, portions of the interview were identified for their possible similarities with each model by the first author. These portions were then grouped together and labeled to assist in organizing all similar quotes together. Additionally, we focused on identifying excerpts where the participants were discussing their experiences specifically within the physics context. After this first phase of coding, the authors discussed the coding until we reached agreement on all the excerpts presented in this paper.

IV. FINDINGS AND DISCUSSION

We present and analyze several examples from the student interviews to demonstrate the utility of models of disability for creating support for Dis. students in the physics community. We believe these examples are useful for mentors in the physics community to consider during future interactions with Dis. students. The following quotes have pauses, filler words and repetitions (e.g., like, uh) removed for easier reading. This paper differs from previous work by highlighting how different models of disability shape different outcomes in similar scenarios. We present findings related to Dis. students’ interactions with physics mentors and Dis. students’ suggestions for improving accommodation access.

A. Interacting with physics mentors

All of the participants discussed interactions with physics mentors. Here, we highlight two interactions, selected because they were clear examples of the affirmative and tragedy models, outlined different ways Dis. physics students may interact with physics mentors, and offered contrasted positive and negative experiences.

Aaron described an encounter with a physics professor that we interpret as aligned with the affirmative model. In response to a prompt about whether there were physics professors who encouraged them, Aaron mentioned that there was one professor that stated that Aaron was a benefit to have in the classroom, and that their autism allowed them
to better succeed. Aaron said, “I really can only think of one professor in my undergrad who's like 'Yeah, this is great. Like your brain works exactly the way it should in physics. And you're pointing out all the things that you should be noticing, and. You are a benefit to have, because your brain doesn't work in the appropriate way.'” This excerpt aligns with the affirmative model due to the professor recognizing the positive that Aaron’s disability brings to the classroom, rather than only the possible negative effects of their disability. Physics mentors can use the affirmative model to support and encourage Dis. students in physics by highlighting the positive opportunities created by the Dis. students’ differences. We encourage instructors to engage with the literature if they are uninformed or unsure about the positive aspects of disability.

On the other hand, Genevieve discussed negative interactions with an instructor when attempting to access accommodations while experiencing generalized anxiety. Genevieve explained that, even after already asking for accommodations (through their local Office of Disability Services), they are forced to remind professors repeatedly while facing possible judgement for asking for their access needs to be met. Genevieve said, “But I have to email the Professor, and then, you know, ask [for] the extension, and sometimes they will have forgotten that I have this accommodation. So, they think I'm just anormal student asking for the accommodation. So that can be very anxiety inducing. And so, a lot of the time, I don't do it. I just like to submit what I have. And submit it on time.” This instance aligns with the tragedy model due to the negative experience, due to their disability, that Genevieve is working through. For Genevieve, the anxiety that they are viewed as someone not needing the accommodation is paralyzing, causing them to miss out on receiving the supports that they need and turn in work that isn’t finished. This quote also is an example of the minority model because of the separation that Genevieve feels between themself and “normal student(s)”. Instructors can help Dis. students by supporting and normalizing the use of accommodations, which makes it easier for students to discuss accommodations with their instructors, ultimately by lowering barriers to access and participation in the physics community. Bustamante, Chini, and Scanlon (2021) provide suggestions about how to create a welcoming environment for Dis. students and suggest that instructors "demonstrate understanding that accommodations promote equity within the classroom." [13]

As instructors and mentors, we can use the varying models of disability to consider possibilities for interacting with students, including ways to encourage Dis. students and potentially reduce the stress they experience in the classroom. In Aaron’s example, their professor’s use of the affirmative model helped encourage Aaron and made them feel included within the classroom. In Genevieve’s case, if the professor operated through the universal model, and gave an indication that they wouldn’t treat Genevieve differently based on their disability status or possible lack thereof, her anxiety may have decreased, and she may have felt more comfortable.

B. Student suggestions about accommodations

In the interview, we asked participants whether they had suggestions for instructors about what they could do or if there was something they wished that their instructors were aware of related to their disability in the classroom. Two examples, from Banner and Genevieve, show examples of the universal model of disability. These quotes were chosen because they distinctly outlined two different, but valid, perspectives of the universal model. Banner describes accommodations while being unaware of which students are disabled, while Genevieve mentions that accommodations can be used for all students, not just those who are disabled.

Banner described that instructors should be ready to make accommodations at any moment, and if unsure, to ask their students what is necessary for them to succeed within the classroom. Banner said, “surely just understanding that accommodations are needed, and then just ask the students what accommodations might be needed. I actually that point would be, you know basically I kind of can't think of what the word is. But basically be prepared to take any accommodations ahead of time.” Banner’s suggestion best falls under the universal model because it exists within the implication that anyone in a classroom could be disabled, so instructors should prepare supports ahead of time, and ask students if they are unsure about what to do. Scanlon et al. [14] identified the most requested accommodations for courses that were taught in emergency remote teaching modalities due to the COVID-19 pandemic. We suggest that instructors inform themselves of the common accommodations used in physics courses and plan for the implementation of those accommodations into their teaching. Additionally, when discussing common pathways to access for Dis. students, Chini and Scanlon [15] suggest instructors “create an ecosystem of supports with a specific focus on transparency by explicitly sharing the resources available through each route in a publicly accessible space, such as a course website or shared group document.”

Similarly, Genevieve states that more generic accommodations would be beneficial to have, not just for themselves, but for students not normally considered Dis. These accommodations wouldn’t be mainly focused on impairment but would also be focused on those who may not be as good in certain academic areas, or those who are first-generation college students. Genevieve said, “I feel like there’s resources that could be good for everybody that I could think of. But, I, there's not anything specific to my disabilities that I wish that I would have.” When asked what specific resources they were considering for everybody, Genevieve continued, “I think it would be good for grad students to have, I guess, for this for the department to have certain types of workshops, like a technical writing workshop, or how to make a presentation for a conference.
like some workshops like this to help students who have, So a lot of people in physics come from families who were in physics, at least that's what I've noticed. And so just kind of some accommodations, I guess, for students. ’’ Genevieve’s suggestion also aligns with the universal model because of the implication that even those who may not be conventionally viewed as disabled (e.g., first-generation students) would benefit from accommodations.

While the final subject between the two was different, as Banner was specific about impairments, while Genevieve had a more general look on accommodations, both describe solutions that we interpret as aligned with the social and universal models. By focusing on accommodations as a solution, both participants imply that changing the social structure of the university would help those with disabilities, which aligns these excerpts with the social model. However, both use the universal model in unique, but equally valid methods. Banner’s statement that the instructor is unaware of which student might need accommodations, and which supports will be necessary ahead of time implies that any student within the classroom could be disabled, and that, for the start of the class, no threshold exists between the disabled and non-disabled within the classroom. Genevieve’s statement goes further by including first-generation college students or graduate students unfamiliar with academia as being able to benefit from accommodations. Just as the universal model states that everyone can be disabled, it also holds that the idea of disability itself is a spectrum, one not just limited to biological differences.

C. Potential model tensions

While we examine the utility of these models of disability for the physics community, it is also imperative that we discuss misinterpretations that may arise through the usage of these models. We want to avoid giving the indication that one model is better than the other and should always be used. Each model has a benefit to its use, and each model can cause harm to Dis. individuals and students if over-relied on. Supported through quotes in our interviews, here we highlight how the affirmative model can be harmful for students, and how the medical model is used by Dis. students to situate their identity.

An example of how the affirmative model can be harmful was demonstrated through Aaron’s experience with one physics professor. This professor ascribed to a belief that if someone was autistic, that they were a genius in math and science. The assumption that all autistic people have extraordinary specific abilities is known in disability fields as ‘savant syndrome’ [16]. Aaron said, “He subscribed to an idea of everyone who’s autistic is, like, they’re either Sheldon Cooper [autistic character from Big Bang Theory], or they’re just not worth it, I guess is the best way to say it. ... it’s one of the terrible cultural effects of the Big Bang theory that, uh, if you are labeled as high functioning, which I don’t believe in those labels [functioning labels are terms used by some medical practitioners and carers to describe the level of support an autistic individual needs to participate in society] but you must clearly be already a genius. And, everything must come easy to you in terms of math and science.”[17] Even though the professor was working through the affirmative model (i.e., through describing their perception that autistic individuals are naturally good at math and science), they shaped the model through ableist stereotypes, comparable to the model minority myth within Asian communities [18], creating an “ideal” disabled individual while denying or discouraging those that don’t fit the savant model. This diminished and discouraged Aaron’s ability as a disabled person, rather than supporting them. Similarly, when using the affirmative model, we must take care not to ignore negative effects that disability can have, such as pain or weakness.

Just as with the affirmative model, the social model does not fully describe Dis. experiences in some situations [5, 19]. In the interview, Banner mentions how they see themselves differently because they lack hearing and this lack is something that sets them apart from others who are not disabled. When asked whether they would view themselves as disabled in a world where “accommodations are natural,” Banner said: “It [disability] actually comes from the practical perspective. I'm still lacking hearing compared to everyone else. So still, there is still something that sets me apart from everyone else.”

This viewpoint is grounded within the critique of the social model because of Banner’s belief that his lack of hearing creates the distinction between himself and those who are not disabled. The social model has its uses in helping provide accommodations for students, but it fails at individual attempts at understanding how disability is grounded within some individuals’ identities. This grounding can also be beneficial for Dis. individuals because it allows for a clear distinction of who might be considered disabled when viewed through the minority model, helping some avoid feelings of impostor syndrome.

V. TAKEAWAYS AND NEXT STEPS

As educators and advisors, we should strive to do our best for our students. We have presented these models as a way for us as mentors to improve our teaching methods, our interactions with our students, and the physics community. This is still a preliminary look at using these models to help represent the interactions that Dis. Students experience as they exist within our community. Our next steps are to continue doing the interviews to allow for a broader range of participant experiences.

ACKNOWLEDGMENTS

This work is supported in part by NSF Grant 1750515.
“The prettiest photos are the ones that have happy people in them”: the use of photovoice in an upper-division physics capstone project course

Kristin A. Oliver 1,2, Victoria Borish 1,2, Bethany R. Wilcox 1, and H. J. Lewandowski 1,2

1Department of Physics, University of Colorado, Boulder, Colorado 80309, USA and
2JILA, National Institute of Standards and Technology and University of Colorado, Boulder, Colorado 80309, USA

Photovoice is a type of participatory action research that aims to enable people to act as recorders and agents of change in their communities. When using the photovoice methodology, participants take photos in response to open-ended prompts and write short captions to accompany their photos. At the end of the photovoice process, participants engage in a focus group where they collectively determine some themes that their photos show, allowing them to co-create the research being done. We implemented the photovoice methodology in a project-based upper-division physics course in which the students partnered with a company in the quantum industry to work on a real-world collaborative project. We present here an example of how photovoice can be used as part of a physics course with a focus on some preliminary results from the students’ focus group. These results demonstrate that the focus group allowed us as researchers to gain new types of information from our students that we may not otherwise have learned, and that the students appreciated the photovoice process, particularly after engaging in the focus group activity.
I. INTRODUCTION AND BACKGROUND

Photovoice, a type of participatory action research methodology [1], is a process that allows participants to represent and enhance their community through the taking of, and reflecting on, photos, thereby empowering them to act as agents of change [2]. This methodology has been used in a variety of contexts including education [3–6], where the student voice is not always heard, even when designing new educational experiences [7].

During the Fall 2022 semester, a new type of course was implemented at the University of Colorado Boulder. This was a quantum industry capstone course called Quantum Forge, in which students worked on a team throughout the entire semester to accomplish a real-world goal for their industry partner. In order to learn about students’ in-the-moment reflections and allow opportunities for student ownership over their experiences in the course [8–10], we chose to implement photovoice as one way to assess student experiences in this course. Furthermore, photovoice allows students a way to provide feedback on the course, giving them influence over the future of Quantum Forge.

A. Photovoice

The photovoice methodology consists of equipping participants with cameras and asking them to document some aspects of their lives or their communities through photos. Participants first take photos in response to an open-ended prompt and write short captions explaining the importance of their photos. Then, in a focus group or interview, participants discuss the photos that they have taken with members of the research team, and participants describe the significance of the photos that they have taken. Finally, photovoice projects frequently end with an exhibition, where the findings of the research can be communicated to individuals, often policymakers, who can affect change within the community [3]. Because photos allow participants to capture in-the-moment thoughts and activities, photovoice allows participants to effectively share their personal expertise, knowledge, and experience with researchers and policy-makers [2].

Photovoice was developed based on theoretical literature on education for critical consciousness, feminist theory, and documentary photography [2]. It has been used in many contexts, beginning with a group of rural women in China [2]. It has also been used in educational settings, such as investigating student identity in a classroom focusing on diversity [4]; the academic and financial experiences of community college students [3]; the health needs of students at Rutgers University [5]; and 7th grade students’ conceptualizations of success [6]. Nevertheless, we are unaware of its use in a physics course up to this point.

B. Course Context

We implemented the photovoice methodology in Quantum Forge, a small, upper-division, project-based capstone course. In this course, during the 2022-2023 school year, students partnered with a company in the quantum industry to accomplish an authentic project involving optimizing heat exchangers for a dilution refrigerator. This course meets the university’s research requirement for graduation, so students can elect to take Quantum Forge in lieu of the traditional advanced lab course or a formal research experience.

The course ran for the first time in the Fall 2022 semester, with eight students participating. All eight students worked on a team together to accomplish this project; this teamwork context is vastly different than what most students have encountered in the past. The intended duration of the project was a total of two 15-week semesters; however, during this iteration several students graduated after the first semester and therefore did not take the course for the second semester. Data from only the first semester are presented here.

The students met with the instructor twice each week and then had designated time to work on their project after one of these weekly meetings. Each week, students were asked to respond to open-ended metacognitive reflection questions, as well as the photovoice prompts. The course included additional activities not directly related to the project such as modules teaching industry skills and visits with representatives from local quantum companies. These activities decreased in frequency as the semester went on so the students could focus on their project. Additionally, while students were not required to participate, they were given the opportunity to participate in interviews with the researchers.

C. Motivation for the use of photovoice in this course

We implemented photovoice in this course as a way for us to gain reflections from the students on how the course was going, their thoughts about the collaborative nature of the course, and their thoughts on the quantum industry. Because this is a new type of course, it was important to get feedback from students and understand the student experience each week rather than simply once or twice during the semester. By allowing the students to create visual images in response to open-ended prompts, they could draw our attention to aspects of their experiences that we might not have otherwise asked about. Students can take photos of moments otherwise inaccessible to anyone other than those that are present, meaning that photovoice allows students, instructors, and researchers access to some thoughts, activities, and reflections that would not be accessible using different methodologies [2].

Also, because this course engaged students in important and authentic work, we wanted the students to participate in a reflective practice that would encourage them to own their experience and have a voice that could shape the course community. Photovoice has the potential to do this by changing what is seen by students, instructors, and researchers as “valuable knowledge” [4].

Finally, photovoice has been suggested as a potential methodology to use within the context of physics courses or
programs [6]. Because we are not aware of photovoice being implemented in the context of a physics course before, we wanted to examine this methodology within this new context to see what we might be able to learn from physics students using this tool.

We, therefore, have two main goals for presenting this work here:
1. To introduce how photovoice can be used in a project-based physics capstone course through an example
2. To present some preliminary results from the student focus group.

II. PHOTOVOICE IN QUANTUM FORGE

Students in the course were asked to respond to a photovoice prompt every week. The prompts were varied and focused on students’ experiences with teamwork, interest in the quantum industry, and engagement with the course overall. Each week, the students submitted a photo and a caption for their photo explaining how it related to the prompt via a Qualtrics survey. Example prompts include:
1. Take a photo of something that motivated you to take this course.
2. Take a photo that represents how your team has decided to divide up tasks so far.
3. Take a photo representing what you find most interesting at the moment about pursuing a career in quantum industry.

At the end of the semester, students were invited to participate in a focus group to discuss and interpret the photos that they had taken throughout the semester. They were encouraged to talk as openly about their experiences as they felt comfortable. Seven of the eight students in the course participated in the focus group activity.

During the focus group, the students viewed the entire collection of photos that they had taken throughout the semester. Students were given the option to not share any of their photos in the focus group, but all students elected to have all of their photos shown. The photos were grouped by which theme (teamwork, quantum industry, or course experience) the photo’s prompt best aligned with. The students were then asked several questions about the photos and about their general experience using photovoice, including:
1. What themes do you notice in these photos?
2. Do the comments other people made about your photos resonate with you? Were there any comments that surprised you?
3. To what extent did participating in the focus group activity change your perception of the photovoice activity as a whole?

The research team then compiled a list of takeaways to share with the course instructor. The students were shown this list and offered the opportunity to make any changes they felt were necessary. This allowed us to complete the photovoice cycle of allowing participants to influence the future design of the course.

III. METHODS OF ANALYSIS

The focus group was recorded using an Owl Labs Meeting Owl 3 [11], so that the researchers could review both video and audio of the entire group. Authors KAO and VB watched the focus group recording and collectively created a content log [12]. The focus group content log was used in two ways. First, themes that students identified in their photos were used to categorize the photos. Details from this analysis will be presented in future work. Second, the content log from the focus group was used to identify 1) important parts of the discussion where students pointed out interesting things that we may not have noticed without the focus group, and 2) some student takeaways from their experience with the photovoice process as a whole. These passages of the focus group were then transcribed in detail and some of them are presented in Sec IV.

IV. RESULTS FROM THE FOCUS GROUP

We present here two takeaways from the focus group as an example of the potential benefits of using photovoice in a project-based physics course. The first is a set of student discussions that brought to our attention things that we may not have noticed without the focus group. The second takeaway is about students’ thoughts about the photovoice process and their experience with it in this course.

A. Topics of interest in photos identified by the students

The students identified several themes that were important to them in the focus group. These themes included the excitement and novelty in quantum industry and the course, money and jobs, the students’ prioritization of each others’ happiness, and a new interpretation of an important location.

First, the students brought up the parallels between the excitement and novelty they saw in their photos about the course and the novelty of the quantum industry as a whole. One student, Owen, stated,

“I was saying some of [the photos] are kind of like the excitement or novelty of a new thing, which, I would agree, I think most of us can say this is the first time we’ve been working on this sort of project before. And just, being brand new to it, it’s all kind of fun and interesting.”

Another student, Jasper, replied,

“And I think the industry kind of reflects that feeling a little bit, too. Like, quantum as an industry prospect is a very new idea, I think, in terms of how long it’s existed. So all these companies... all these other people, a big part of what I think motivates their work is the excitement of the novelty of it, and so, I guess that’s part of what makes quantum industry exciting but also maybe dangerous as a career.”
Together, these two students identified that excitement and novelty were a part of their experience with the course and their expectations about the quantum industry. This seems to indicate that the students are seeing the course as a parallel to the quantum industry; they are identifying that their work in the course is novel and exciting in a similar way to that of the industry as a whole. While we have noticed students’ excitement about the course in other places, the focus group gave students an opportunity to share with us the way this feeling paralleled their feelings about quantum industry.

Next, we present an instance where one student was surprised that more of the photos did not center around money and jobs. We highlight here the discussion that followed about how students were viewing the role of money and jobs in relation to this course. One student, Charlie, said,

“I’m surprised more people didn’t take pictures about getting a job. I took the pictures with the money in them and a big reason I took this class is to be able to get a job after college... I’m just surprised that. I guess, other people didn’t take it for the same reason.”

A second student, Reese, clarified that money was an important motivating factor for them, but that their happiness was more important than money.

“I would say that it’s a motivating factor but if I don’t like it then it doesn’t really matter. I could be making like two hundred thousand dollars a year but if I’m miserable then that means, like, almost little to me because I won’t be happy.”

Charlie then replied, saying,

“Yeah, yeah, I agree with you, I don’t wanna make it seem, like, all about the money. I was thinking, it was just– getting a job is, like, necessary for me to be able to do the things I like, you know?”

Finally, Jasper added that they don’t think there’s anything wrong with “saying that you want to have a life of, like, stability... I don’t think there’s anything, like, immoral about wanting that.”

These three students are looking at the role of money and jobs in several different ways, and none of these unique perspectives would have been captured given the photos and captions alone. The focus group was able to highlight the perspectives of students who did not address this theme in their photos and captions, and provided additional nuance to those perspectives that did appear in the photos and captions. The first student volunteered their motivation for taking the course in response to the photos presented, which opened up space for the other two students to add their perspectives on money and jobs in relation to this course.

Another student perspective that we gained from the focus group was about a common type of photo that appeared in their photovoice responses. During the focus group, many students brought up the fact that the whiteboard in their lab space was photographed frequently. We had noticed this as well, but the students added some nuance to the way we thought about these photos. Initially, the students claimed that the whiteboard photos were a consequence of “laziness,” with Owen stating, “I think there’s a very obvious ‘I was lazy this week’ sort of picture, which happens to be the whiteboard this time around.”

Later on, however, another student, Stella, mentioned that there might be something deeper being represented in the whiteboard photos. Stella states,

“I do like that, um, the whiteboard ones are all completely different prompts. It– it’s kinda funny. Like, they’re– they all have sort of like that vibe of teamwork, but every single prompt of the whiteboard photos is different. Which I think just goes to show that, um, I’m not sure if we took pictures of it because it’s... immediately what we go to when we think of this class is ‘oh, the whiteboard. Cause that’s where I’m with everyone.’ Um, I don’t know, I just think it’s kinda nice.”

Reese then agrees that the whiteboard photos all fit very well with the prompts, reiterating that their experience with teamwork largely happens around the whiteboard. Owen then mentions that of all of the equipment available to them, they use the whiteboard most and that it has been “doing the heavy lifting.”

This discussion led us to determine that the many photos students took of the whiteboard had more significance to them than we, or they, initially thought. The focus group, therefore, allowed us to identify a location that was important to the students and their teamwork experience that we may not have otherwise understood in this way.

Finally, the students at the focus group brought up the fact that the happiness of the group was a priority to them. Owen said,

“I do think it’s interesting how my photos all have pictures of people. It kind of speaks to the way that I’ve approached the project. I absolutely care most about the people. The progress doesn’t matter so much to me. If you guys are unhappy then I don’t care how far along we are, you know?... I think the prettiest photos are the ones that have happy people in them.”

This student identified the fact that they had taken photos mostly of people and tied it to the fact that they were invested in the group’s happiness and cared most about the people, which is a connection that is not immediately visible from the photos or the associated captions. This sentiment was echoed by other students who agreed that the happiness of the group was their priority. Without the focus group, we would not have had the opportunity to learn how this student cared about the well being of their group as a whole.
B. Student reflections about the photovoice process

At the end of the focus group, we gave students the opportunity to provide feedback on their experience with the photovoice process. Overall, the students expressed that the experience of engaging with photovoice was valuable and helped them realize things about their teammates and the course.

For instance, the students pointed out that the photovoice process allowed them to realize that they all had a similar experience with the course. During the focus group Owen said, “We were definitely thinking about a lot of similar things and this really brought that to life. I wouldn’t have otherwise known.” Stella agreed, “I do think it’s really nice to see how everyone is, like, we’re all on the same boat. And it’s just nice to see what everyone took photos of, honestly.”

Both of these quotes show that students felt that the photovoice process allowed them to understand that their feelings about the project were shared by others, indicating that photovoice can bring students together over shared experiences.

The students also emphasized that the focus group experience itself was incredibly important to their experience of the photovoice process. For instance, when asked how the focus group impacted their perception of the photovoice experience, Jasper replied, “It really brought it together, I feel like. It’s the ‘voice’.”

As students continued to talk about the impact of the focus group, Stella said, “I agree with everyone, it brings it all together. Because some weeks I was like, ugh, I have to do the photovoice!... and I have to do my reflection, and I have to submit, like, my lab scan, and all of this other stuff on the same Friday, but then, like, seeing it now, I definitely think it was more valuable than I realized at the time.”

These students indicated to us that the focus group tied together their experience with photovoice as a whole, allowing them to see the value in the activity and to bring their voice to the table.

Finally, the focus group allowed us to hear from students that reflection in the form of photovoice was more memorable than other forms of reflection. For instance, Reese said, “Yeah, I would also say it’s also a lot easier to remember when these were taking place and what not, so if I were just, like for the [reflection questions], I probably could not tell you what day of the week, what day of the month, or semester I wrote those. But this I could give you a rough estimate of, ‘oh that was like the first two weeks,’ so it’s like, yeah, you can better track your progress that way.”

This demonstrates that reflection in the form of photovoice helped students engage in self-reflection in a different way than traditional forms of self-reflection, allowing them to reflect back on their photos at the end of the semester and remember specific moments about their journey throughout the course.

V. CONCLUSIONS

The implementation of the photovoice process in Quantum Forge yielded valuable information for both researchers and the broader physics education community. These preliminary results from the focus group allow us to share some of the ways in which photovoice can benefit both of these groups.

One way that the photovoice methodology benefited us as researchers is that we came away from the focus group activity with a list of themes that will guide our future analysis of photovoice photos and captions, interview data, and reflection questions. Through their participation in the focus group, the students were able to contribute to the research process by telling us which themes they saw in their own work that were important to them. As mentioned in Sec IV, many of these themes would not have been noticed or interpreted in the same way without student input.

Furthermore, the focus group allowed us as researchers to better understand how to improve the photovoice experience for students in the following semester of the course. The students indicated that they would appreciate having two focus groups in the next semester of the course.

In addition to researchers, instructors can also benefit from the data produced by photovoice. These data are potentially different than other forms of data typically collected by instructors and education researchers. Furthermore, these data are guided by what students find most important, allowing us to have insight into what is valuable to our students that we might not otherwise know to ask about.

Finally, from the student perspective, photovoice has the potential to provide a positive experience, allowing them to share their ideas and experiences with researchers and with one another, and giving them an opportunity to engage in self-reflection that is memorable and meaningful.

While photovoice has given us the opportunity to learn a great deal from the students in Quantum Forge, there are some significant limitations to implementing photovoice in a class setting. For instance, engaging students in a focus group activity may be challenging in a larger class, as well as getting students to buy into the process.

Although photovoice may not be a viable option in all physics courses, we have demonstrated that it may be a fruitful research methodology in some situations. Especially in classes where student feedback is particularly important, photovoice may allow researchers and practitioners access to unique perspectives and allow students to affect change in their educational experiences.

VI. ACKNOWLEDGEMENTS

We would like to acknowledge the PER group at the University of Colorado for providing us with their feedback on this work. We would also like to thank the students in Quantum Forge for sharing their experiences with us, as well as the instructor for his support throughout this endeavor. This work is supported by NSF QLCI Award OMA 2016244.


259
Alignment between student epistemological views and experiences with course structures in introductory physics: A case study

Ellen Ouellette, Sarat Lewsirirat, Ryan Biju Sebastian, and Morten Lundsgaard
Department of Physics, University of Illinois Urbana-Champaign,
1110 W Green St Loomis Laboratory, Urbana, IL 61801

Christina Krist
Department of Curriculum & Instruction, University of Illinois Urbana-Champaign,
1310 S Sixth St Education Building, Champaign, IL 61820

Eric Kuo
Departments of Physics and Curriculum & Instruction, University of Illinois Urbana-Champaign,
1110 W Green St Loomis Laboratory, Urbana, IL 61801

Designing physics courses that support students’ activation and development of expert-like physics epistemologies is a significant goal of Physics Education Research. However, very little research has focused on how physics students’ interactions with course structures resonate with different epistemological views. As part of a course redesign effort to increase student success in introductory physics, we interviewed introductory physics students about their experiences with course structures and their learning and belonging beliefs. We present here a case from this broader data corpus in which a student, Robyn, discusses his epistemological views of physics problem solving and his experiences with physics lectures, office hours, and discussion sections. We find that Robyn’s physics epistemology manifests consistently across his interactions with each of these different course structures, suggesting a possible resonance between students’ beliefs and their experiences with course structures and the value of further investigation into the potential merits of comprehensive course design.
I. INTRODUCTION

Physics classes impart not only content knowledge but also messages about the nature of physics knowledge and learning. Students’ personal epistemologies of physics have been classified along different dimensions, such as beliefs about the structure of physics knowledge, beliefs about the content of physics knowledge, and beliefs about learning physics [1, 2]. Beliefs along each of these dimensions range from more expert-like to more novice-like in character. For example, an expert-like belief about learning physics might be that learning physics is about working to develop one’s understanding and intuition of concepts, while a novice-like belief about learning physics might be that learning physics is about memorizing information received from an authority, like an instructor or textbook.

Helping students activate and develop epistemological views that can support their long-term success has emerged as an explicit learning goal in Physics Education Research. Expert-like epistemologies of physics have been associated with students’ deeper engagement with content knowledge [3, 4] as well as better course performance and higher rates of retention [5]. Unfortunately, traditional physics instruction has often had the opposite effect; most studies which survey university physics students’ beliefs before and after instruction find that students leave physics courses with less expert-like beliefs about the nature of physics knowledge and learning than they initially reported [6]. However, a few specialized approaches, such as an explicit focus on epistemological development [7] and modeling instruction [8], have produced positive shifts in students’ epistemological stances, as measured by the CLASS and MPEX surveys.

Compared to the many pre-post survey change studies, very little research has focused on how specific course features in physics courses and students’ experiences of them resonate with different epistemologies of physics. However, a few studies which interviewed STEM students about their experiences with course structures and their epistemological views have found evidence to suggest a relationship between the two. For example, high school mathematics students’ interactions with didactic or collaborative learning environments have been linked to their adoption of views that mathematics knowledge is definite and received from authority or constructed through independent thinking respectively [9]. Additionally, undergraduate engineering students’ interactions with assessment and grading policies which do not reward deep understanding have similarly been linked to their views that this type of thinking is not valued in the engineering program [10]. Further study of how students experience and interact with physics course structures at the college level can help inform future instructional design efforts.

In this paper, we present a case of an introductory physics student, Robyn (pseudonym), discussing his epistemology of physics learning and problem solving as well as his interactions with course structures. Our analysis addresses the following research question: how can students’ epistemological views become evident in their interactions with course structures? We will illustrate that Robyn’s epistemology of physics manifests in consistent ways across his interactions with a variety of different course structures. This finding suggests the value of further research on how course structures can be designed coherently such that they resonate more strongly with the epistemologies we desire for our students.

II. METHODS

In the Fall of 2022, as part of a larger course redesign effort and investigation into the interaction between students’ experiences of course structures and their beliefs related to learning and belonging in physics, we conducted semi-structured interviews with students enrolled in Physics 100, an introductory physics course at a large, public, midwestern university. Enrollment in Physics 100 consists primarily of freshman engineering majors. The course provides these students with a one-semester introduction to kinematics and dynamics to prepare them to take Physics 211, a calculus-based introductory mechanics course which is required by many major programs, including all engineering majors. Weekly course meetings consist of one lecture (preceded by an at-home video pre-lecture) that incorporates peer instruction-style questions and one discussion section where students solve problems in small groups. Students also complete two homework assignments each week and are encouraged to attend office hours to receive help from peers and instructors with the homework questions.

Semi-structured interviews followed a protocol designed to probe both students’ experiences with different Physics 100 course structures and their beliefs about learning and themselves as learners. Interviewers selectively diverged from the pre-planned protocol questions to pursue experiences and topics that were salient for the students, particularly if they were related to students’ learning beliefs. Students were asked both open-ended questions about their Physics 100 experience as a whole, the challenges they faced, and their views on learning physics as well as questions about the usefulness of specific course features. Students’ responses to a pre-interview survey on these topics were used to guide parts of the interview. All interviews were video and audio recorded, and the recordings were transcribed and discursive turns numbered. Interesting segments related to the project research questions were flagged for subsequent, more in-depth analysis which included viewings and discussion in larger research team meetings.

Robyn (whose preferred pronouns are he/him) was interviewed in week 12 of his first semester of college. At the time of his interview, Robyn was enrolled in the engineering college as “undeclared,” meaning he had not yet selected a specific engineering program, though he reported interest in pursuing systems engineering on his pre-interview survey. During the interview, Robyn described having had a very negative view of physics at the start of the semester based on his
experiences with physics in high school. However, at the time of the interview, Robyn claimed that his very positive experience in Physics 100 had altered this view to the point where he was considering selecting physics as his major.

Robyn’s interview was selected for this case study because he was especially articulate about his epistemological views in physics and his views on Physics 100 course structures. Consequently, Robyn represents an exceptional rather than representative case from the data corpus, which we use to explore theoretical ideas of how students’ beliefs may be connected to their experiences with course structures.

III. RESULTS

A. Robyn’s epistemological view of problem solving in physics: general application of a few key concepts

We will start by characterizing Robyn’s general epistemological view of problem solving in physics. Early in the interview, Robyn indicates that learning physics has been challenging. The segment of transcript below begins with a followup from the interviewer, prompting Robyn to elaborate on those challenges. Robyn proceeds to explain that he perceives a difference in problem solving in physics versus in other disciplines.

9 I: When you say [physics] is challenging, what aspect of it is challenging?

10 R: Just understanding physics... it requires a completely different way of thinking than other STEM courses like calculus, chemistry, biology. In high school, I had a lot of chemistry, a lot of biology, I mean the standard calculus, precalculus, algebra two courses. All of those, they’re more like just, it’s a lot of computational work where you’re given numbers, and you’re given one problem, and you’re just told to solve it using one strategy that you learned how to solve in class or lecture. But physics, it requires thinking in a different way. Rather than thinking about how to get the, like, just your one goal is to get the answer, it’s thinking about the many different ways there are to get the answer, and there’s more I’d say theory versus just a plain, just a simple question. And, yeah, just adapting to that way of thinking has been challenging, but really, I feel like it’s three big concepts. You have, everything is based around Newton’s laws, so I feel like once I understood those, and I’m still learning how to understand those, but I feel like that understanding has been key to my success as the course has gone on. So first understanding that was challenging, but now it’s gotten easier, I think. And you learn how to think about the questions you’re given, and you learn how, like, to write out equations and draw diagrams is just the general strategy for any physics problem you’re given.

Epistemologically, Robyn’s description of physics problem solving focuses on general, underlying concepts rather than memorizing and applying prescribed procedures. Robyn describes physics problem solving as focused on theory and key concepts, such as Newton’s laws. He then contrasts this experience with problem solving in other courses, which can be summarized as plug-and-chug: he is assigned mainly “computational work where you’re given numbers” and taught one prescribed strategy to apply to each type of problem.

Next, we will examine Robyn’s reported experiences in three different parts of the course: lecture, office hours, and discussion sections. We focus on these course structures because Robyn is especially descriptive about how they support his learning in ways that we will argue are aligned with his epistemology of physics.

B. Robyn’s experiences with course structures: valuing understanding through independent thinking

In recounting his experiences with three different course structures - lecture, office hours and discussion sections - Robyn expresses repeatedly that he values opportunities to develop a deep understanding of concepts rather than memorize and apply procedures, which we contend is aligned with the epistemology of physics learning Robyn describes earlier in the interview.

In the next section of the interview, the interviewer asks Robyn what parts of the course have been most helpful. Robyn starts by explaining that answering Peer Instruction-style questions in lecture with clickers is useful because they encourage students to think deeply about problems and their solutions rather than just accepting correct answers provided by instructors.

19 I: What aspects of the course do you find most helpful?

20 R: ... the iClicker. We use that in lecture a lot, and I kind of compare it to the use in chemistry, and even though they’re the same device, I use it in a completely different way in physics. And our results in physics, we actually discuss them, and [the] Professor, he goes over why the wrong answer is wrong versus just why the right answer is right, which is what we do in chemistry. It’s like, 50% of students put answer A, 50% put D, and D’s the right one. In chemistry, our professor just goes "Ok, here’s why D is right." But in physics, we talk about why A is wrong, and then he asks for student opinions. So if I put A, and it’s wrong, he asks me to... defend why I thought A is right, and then he asks a student who put D to defend why D is right. And then he asks us to talk with our table again and re-answer the question. So the iClicker questions encourage more collaboration and more thinking, which leads to better results and more success in the class.

21 I: Got it, ok. So, is there... something in Physics 100 that you would like to see implemented in other courses?
22 R: ... I know the iClicker is implemented in other courses, but I would like to see the way that we use the results of the iClicker to be used in other courses, like more discussion on why something is wrong. Because if I put a wrong answer, obviously I thought it was right, so I want to know why it's wrong, versus why the people who put the right answer got it right. Because you don't necessarily gain anything from just seeing the right answer, in my opinion.

Overall, Robyn describes two valuable elements of iClicker questions in physics lecture that go beyond just learning the correct answer and why it's correct. One is discussion of the questions with peers, which encourages students' independent thinking and deeper understanding. The second element is an intentionally designed addition to Physics 100 classroom discussions: explicit discussion of why seemingly sensible, incorrect answers are wrong.

The benefit Robyn ascribes to this lecture activity highlights an epistemological stance towards learning physics aligned with his general view of physics problem solving. Here, Robyn values opportunities to practice his thinking with the underlying ideas, not just learning the correct answer. He even goes as far as to say "you don't necessarily gain anything from just seeing the right answer in my opinion." This epistemological view of learning fits with the epistemological category "independence in learning physics" [1, 2]. It also fits with the epistemology of problem solving as application of a few general concepts rather than applying learned procedures that Robyn described previously, since applying concepts to physics problems requires more independent thinking and deeper understanding than applying learned procedures.

Later, the interviewer prompts Robyn to discuss the skills he feels he's developed as a consequence of taking Physics 100, and Robyn's answer, which focuses on office hours, again reflects his belief that learning physics requires understanding concepts rather than memorizing procedures.

33 I: Could you give me some examples of the skills that you have gained from Physics 100?

34 R: ... Another skill that I find myself using a lot is inquiry... Sometimes, like in office hours last week, [the] Professor explained something to me, and it really made no sense, but I was able to ask him to explain it another way, and it made perfect sense. So the same idea but just explained two different ways. So it's really enhanced my skills of asking questions and being comfortable asking for another way of solving a problem rather than just, like, having one way to solve it and not understanding it but kind of just going with it.

In describing his experience with office hours, Robyn once more expresses a reluctance to accept and internalize procedures without deeper understanding. Robyn's belief in the utility of "inquiry" as a learning strategy seems rooted in his view that "just like having one way to solve it and not understanding it but kind of just going with it" is not the way to learn physics, as supported by his statements regarding problem solving in Turn 10 as well. Robyn's description of his experience in office hours is also very similar to the statement he makes about the value of discussing the wrong answers to iClicker questions, where Robyn was similarly dissatisfied with the idea of having to accept a solution approach that he didn't fully understand.

Later, Robyn is presented with a hypothetical scenario in which he is asked to reflect on his experiences in discussion sections and given a choice between joining a group where one person knows the right answers and leads the rest of the group in solving the problem (Samira's group) and another group where the students are all more uncertain but work together to arrive at a solution (Meena's group). Robyn has just explained that his experience in discussion sections has been very similar to Meena's in that his group works together to construct answers to problems and that he prefers this method over relying on the guidance of a single group member. The interviewer follows up by prompting Robyn to consider the possible merits of Samira's group, and Robyn's response is similar to the interactions he describes with the previous two course structures and aligned with his epistemological views on physics problem solving.

123 I: ... it seems like Samira's group works very efficiently and completes the task quickly. Does that seem like a good thing to you as well?

124 R: I mean, yes, efficiency is always good, especially when you don't have that much time, but it says (reading from and pointing to print out of scenario), "one person who always knows the right answer, so we pretty much follow her lead." So I don't know if that's necessarily, like if you know the right answer, I'm just going along with you even if I don't understand it. So that's definitely not effective, and... I'd rather know how to do a problem than have the work done on time but not know how to do the problem.

When asked to choose between completing a task and understanding concepts in the context of discussion sections, Robyn values the experience he's had collaborating on problems with his group members and explains that "I'd rather know how to do a problem than have the work done on time but not know how to do the problem." Once more, this statement is extremely similar to Robyn's reluctance to accept a solution from an instructor that he doesn't understand in either office hours or lecture, and to the more general beliefs Robyn expresses early in the interview about learning physics as learning to apply concepts and understand the underlying theory rather than prescribed procedures.

In summary, Robyn's epistemology of physics - that learning to apply and understand concepts is more valuable than memorizing and applying procedures - is evident in his assessment of physics problem solving as well as his descriptions of his experiences in lecture, office hours and discussion sections. Robyn repeatedly expresses dissatisfaction with accepting solutions that he does not fully understand and seeks out opportunities to use course structures to help him refine
his physical reasoning, illustrating the connection between his epistemology and interactions with course structures.

FIG. 1. Robyn’s epistemological views and experiences with iClicker in lecture, office hours, and discussion sections. The up-down arrows indicate hypothesized mutually reinforcing relationship between epistemological beliefs and experiences with course structures, which can be clarified through future research.

IV. DISCUSSION

This case study provides an example of how a student’s epistemological views can be aligned with their perceptions of their interactions with course structures, such as lectures, office hours, and discussion sections. Figure 1 summarizes Robyn’s espoused epistemological views and experiences with course structures and indicates a hypothesized cyclic reinforcement loop between students’ epistemological views and their experiences with course structures. In Robyn’s case, interactions with course structures resonated with a sophisticated view of physics learning. However, not all of the students in our study reported interacting with these course structures in the same way as Robyn, and their accounts of the epistemological aspects of their experiences are generally less explicit and consistent. Our future work will attempt to apply a similar analysis to other students’ reports of their experiences to investigate other possible resonances between students’ interactions with course structures and their epistemological views. Additionally, we plan to survey students’ epistemologies as well as collect observational data of course activities and compare these findings with epistemological aspects of students’ reports of their course experiences. We hope this analysis will help us better understand the details of our hypothesized resonance loop and advance our instructional design principles for supporting students’ development of expert-like epistemologies.

The alignment in Robyn’s experiences with different course structures suggests another key issue for future research: the importance of aligned messaging between multiple course structures. Physics instructional design often focuses on deliberate efforts to develop students’ knowledge and beliefs, such as Hammer and Elby’s instructional approaches to help students become aware of and refine their everyday thinking [11]. However, we propose that other course structures can inadvertently reinforce or clash with the epistemological goals of these explicit instructional efforts. For instance, grading that focuses on the correctness of numerical calculations may support epistemological views that learning physics means learning to apply formulas for computation, thereby undermining the effectiveness of other efforts to develop students’ expert-like epistemologies. A better understanding of how different course structures can resonate with different epistemological views could help explain why pre-post survey data show that many piece-wise PER-based instructional reforms fail to positively develop students’ epistemological views and why courses that produce positive pre-post epistemology/attitude survey shifts, such as an explicit epistemological curriculum or modeling instruction, shift multiple course structures to align with the learning goals.

Similarly, our findings also suggest that it may be fruitful in future work to attend particularly to the alignment between the epistemological messages of both instructor practices and course structures. In Robyn’s accounts of his experience in lecture and in office hours, he attends particularly to the practices of the instructor: in lecture to their choice to discuss wrong answers when using iClicker questions and in office hours to their willingness and ability to explain a concept multiple ways. In our analysis, we considered these practices as part of the course structures in which they occurred, but future work can break down the ways in which instructor practices and course structures can reinforce each other to support students’ development of expert-like epistemologies.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant Nos. 2100081 and 2235516.
“Which has more energy?” - An example of responsive teaching in university physics

Jon C. Owen (he/him/his)
Physics Department, Seattle Pacific University, 3307 3rd Ave W, Seattle, Washington, 98119

Lisa M. Goodhew (she/her/hers), Brynna Hansen (she/her/hers)
Physics Department, Seattle Pacific University, 3307 3rd Ave W, Seattle, Washington, 98119

In responsive teaching, instructors seek to understand and pursue the substance of students’ thinking by foregrounding connections between students’ ideas and disciplinary understandings. Education research literature suggests that responsive teaching benefits student learning, yet is also difficult to implement in fast-paced science courses, including university physics courses. This may be one reason there are few examples of responsive teaching at the university level. We share an example of responsive teaching from a small group conversation about heat and temperature in an introductory, calculus-based university physics course. In this example, an instructor proposes a thought experiment that takes up and advances students’ thinking about heat and temperature. This example illustrates that responsive teaching is possible in university-level courses, and suggests that one way university science instructors can bridge students’ thinking and sophisticated content learning goals is by posing carefully-selected “thought experiments” that target the connections between students’ thinking and scientific models.
I. INTRODUCTION

Responsive teaching is an instructional approach that is “grounded in an empirically and theoretically supported expectation that students’ intuitive thinking about science is productive and resourceful,” [1]. Science instructors enact responsive teaching by noticing the “seeds of science” in their students’ thinking [2] and then leveraging connections between student ideas and canonical science models to support the development of students’ understanding [3,4]. Therefore, the curriculum and learning goals emerge (at least to a certain extent) in real time, as students engage with the material an instructor presents to provide a basis for exploration [5]. Literature on responsive teaching suggests that this teaching approach has the potential to promote students’ growth as scientists more effectively than more traditional approaches. A number of studies have shown that responsive teaching practices improve students’ conceptual understanding, promote students’ agency in the classroom, and engage students in authentic scientific practices [1,2,6–10]. Here we present a short example of responsive teaching from a university physics class session. This comes from a classroom that used a worksheet that was specifically designed to be more open-ended than typical instructional materials for university-level physics. This worksheet – an ACORN Physics Tutorial on heat, temperature, and thermal energy [11] – provided an optimal context for enacting responsive teaching. This case study examines what responsive teaching strategies may look like amidst the inherent instructional challenges of an active, student-centered introductory physics course.

Though responsive teaching can promote students’ growth as scientists in multiple ways, there are very few examples of responsive teaching in university science classrooms [12]. Published examples of responsive teaching in STEM classrooms predominantly come from K-12 settings [1–3,13,14], which may suggest that responsive teaching is more appropriate for pre-college students. These examples depict teachers spending significant time clarifying and extending their understanding of students’ thinking, then adapting their instruction and curriculum to build on students’ fruitful ideas. In these examples, student ideas take up much of the “airtime” in discussions; science principles and correct answers take up less discursive space. Several of these examples articulate a tension between pursuing students’ science ideas and guiding students to the canonical result, model, or understanding [2,13,15]. While a commitment to responsive teaching does not mean that canonical scientific understandings are unimportant [16,17], it may mean that instruction follows a circuitous, dynamic path, where some goals get more time than others, and a broader range of intellectual trajectories are planned for.

University physics courses are typically fast-paced and mathematically rigorous, and the content expectations for a university-level introductory physics course pose a particular challenge to responsive teaching. Knight’s calculus-based Physics for Scientists and Engineers text [18], the text used in the course studied here, includes approximately 30 chapters that are covered in three 10-week quarters or two 15-week semesters – there is little “wiggle room” in the course schedule. In other words, the inherent tensions of responsive teaching are likely to be particularly challenging in a university-level physics course. Yet, Robertson et al. give an example of responsive teaching in an introductory physics lecture course that suggests that it is possible, at least on short timescales [15].

In this paper, we add to the illustrative case given by Robertson et al., examining what we claim is a responsive teaching interaction in a calculus-based university physics course. Our example shows the emergent and dynamic interaction of responsive teaching in a new context – the physics topic and the course structure are both different from the example in [15]. In this example, a question posed by an instructor builds on students’ productive reasoning while also building towards the content goals of introductory physics. The instructor’s question closely relates to many of the ideas the students discussed earlier in the class session and explores those ideas further in a “thought experiment.” We argue that this case illustrates that responsive teaching is possible in university-level courses with rigorous content learning goals and is an example of how such responsive teaching interactions might look.

II. INSTRUCTIONAL CONTEXT

The case we share is from a small-group discussion that took place in a calculus-based introductory physics course at a small (<5000) liberal arts university in the Pacific Northwest United States. This course primarily serves students majoring in engineering, computer science, physics, chemistry, and biology. The racial and gender demographics of the population served by this course is as follows: 48% Female, 52% Male1; 7% are international students; students who are residents of the U.S. are 44% white, 17.8% Asian, 8.3% Black or African American, 0.3% Hawaiian Native/Pacific Islander, 14.9% Hispanic of any race, 8.1% Two or more races; we do not know how the particular course studied may differ from this larger population. The class in which our data was collected was the second of a three-quarter introductory physics sequence, composed of approximately 30 students, taught by one experienced...

---

1 The university reports gender demographics as male/female; however, we do not assume this accurately represents the spectrum of identities held by students in this population.
faculty member and supported by two undergraduate Learning Assistants. Another faculty member was present in the classroom when video-recording occurred. A significant portion of class time was dedicated to small-group work scaffolded by worksheets, including Tutorials in Introductory Physics [19] and other materials. Students typically worked in table groups of 3-5.

Several student groups were video-recorded during two class sessions in which they worked through an Attending to Conceptual Resources IN (ACORN) Physics Tutorial on concepts of heat, temperature, and thermal energy [11]. The worksheet asked students to explain a set of experiments involving heat transfer by conduction. For example:

“Two identical metal blocks are sitting on a table. One is hot and one is cold. The blocks are placed in contact with one another and put into an insulated box. After several minutes, the blocks are the same temperature as one another. (1) Why does this happen? (2) Is the final temperature of the blocks halfway between their initial temperatures, between their initial temperatures but not necessarily halfway, or not necessarily between their initial temperatures?

Then the worksheet directed students to articulate, apply, and refine their ideas about thermal phenomena, ultimately guiding students to generate a set of rules that predict and explain heat transfer by conduction. This process of addressing a scenario and constructing a set of rules is iterative and the students’ rules are expected to change and grow over the course of the worksheet. In our observations of the class using this worksheet, we notice that the goal of articulating rules about thermal phenomena guided the direction of students’ conversations and instructor moves more than the specific scenarios presented did. Students spent significant time articulating, refining, and making sense of rules; in doing so, conversations often diverged from the worksheet scenarios to other thermal phenomena.

III. THEORETICAL FRAMEWORK AND METHODS OF ANALYSIS

This paper presents a case study of a small-scale responsive teaching interaction in calculus-based university physics [20,21]. For this study we video recorded all groups for which we had consent from all students. We selected the focal episode of this paper from a larger corpus of video recordings that captured four groups of students for the entire class period on the day in which the ACORN Physics heat and temperature worksheet was used in the classroom described above. We began our analytic process with broad thematic questions about how instructor interactions support students’ progress through the ACORN Physics worksheet, and used an inductive approach to refine our research questions and claims [22]. This process led us to highlight the focal episode of this paper as an instance of rich, extended instructor-student interaction. We iteratively viewed and discussed the video of this conversation, discussing possible interpretations of and claims from this episode [23]. Our interpretation was guided by the marks of responsive teaching articulated by Robertson, et al. [15]:

a. Foregrounding the substance of students’ ideas;

b. Recognizing the disciplinary connections within students’ ideas;

c. Taking up and pursuing the substance of students’ thinking.

We applied principles (a) and (b) as a first filter, marking this as an interaction in which students’ science ideas were apparent to us, and in which an instructor responded to the disciplinary content of those ideas. After identifying this candidate episode, we analyzed the transcript closely, looking for evidence of instruction that took up, pursued further, or built on the physics content of students’ ideas that was apparent to us and that was taken up by the instructor. In the following section, we unpack how this episode fulfills these criteria to illustrate responsive teaching in a university-level physics classroom.

FIG 1: Image of the oil and water scenario from the Energy Forms and Changes Simulation by PhET Interactive Simulations, University of Colorado Boulder, licensed under CC-BY-4.0 (https://phet.colorado.edu).

IV. ANALYSIS

This analysis focuses on a conversation between three students, pseudonymed Sam, Stephanie, and Selena. At various points in the class period, the students discuss their thinking with Professor Pete. Although the focal episode begins during the middle of the activity, we believe it is important to describe what happened before the focal episode as it provides an important context.

The class session begins with a review of energy conservation, and Pete presents an energy conservation equation that includes a term for heat transfer. Pete introduces the heat transfer equation $Q = mc\Delta T$. He explains that $Q$ is the heat transferred to a substance, which depends on its mass, change in temperature, and its specific heat capacity $c$. Pete shows a table of the specific heat capacity values for various materials. Then, Pete presents a scenario for the class to discuss in small groups: “If we add the same amount of heat to the water and oil, [so] $Q$ is the same in both of them, will the temperature change be the same? If not, which one is a higher change of temperature?”
As he presents this question to the class, Pete shows the scenario in the PhET Energy Forms and Changes simulation (Fig. 1) [24], which the students used in a pre-class assignment. The simulation depicts energy units and thermometers that qualitatively show the change in energy and the change in temperature of water, oil, brick, or aluminum when they are heated or cooled.

Sam, Stephanie, and Selena begin to discuss the oil and water scenario. Initially, Sam predicts that “the oil heats up faster than the water.” Selena agrees, stating that “the oil needs less energy to be at that temperature.” The group discusses the equation $Q = mc\Delta t$ and how it applies to the oil and water scenario for several minutes. At the end of their discussion, Pete checks on the group and asks about their prediction: which liquid has a higher temperature, and which gains more heat? The group shares that the oil will have a higher temperature, but each liquid has the same heat added. Pete affirms their thinking and adds, “so heat and temperature are not the same.” We infer that distinguishing between heat and temperature is an important learning goal for Pete, which we see resurface later in our focal interaction.

Following this conversation, the class is instructed to begin the worksheet. After considering the first question in the worksheet (the question given in section II above), the group agrees that the final temperature of the two identical blocks must be halfway between the two initial temperatures, or the “average.” They explain that “the transfer of energy goes from high to low,” and write this down as a rule. They also express uncertainty that the final temperature would always be the “average.” Stephanie says, “That bothers me. Like there has to be one where it’s not the average, but I guess that has to do with like energy, not staying in the system that causes that.” The group discusses the same ideas for a few more minutes without reaching a confident resolution.

Our focal episode begins just afterward, when Pete comes up to the table and asks about the group’s progress:

**Pete:** Do you have any rules yet?

**Sam:** Oh. I haven’t even gone back yet.

**Selena:** Uh, energy likes to go from high concentration to low concentration.

Here, Pete opens the conversation with a question about the group’s progress toward the goal of the worksheet. We interpret this instructional move as foregrounding students’ ideas because it is an open-ended question that invites the students to share a summary or important idea. While Selena’s brief response does not convey all of the ideas that the group used in the preceding conversation, it does give Pete insight into their thinking that he uses to dig deeper into the group’s understanding. Pete revoices Selena’s rule, adding particular emphasis to the energy idea:

**Pete:** So the energy goes from high concentration to low concentration?

**Stephanie:** Mhm.

**Pete:** Um, can I ask a question?  

---

2 Italics indicate the speaker’s emphasis.  

---

**Stephanie:** Yeah.

**Pete:** Oil, say at 30 °C, water at 10 °C. Which has more energy?

Revoicing students’ ideas, as Pete does in the first line in the snippet of conversation above, is affirmed in the literature as a responsive instructional move [5,25,26]. We interpret this instructional move as both drawing the students’ attention to the energy idea (particularly given Pete’s emphasis), and solidifying Selena’s meaning for the group, as Stephanie agrees in response. This instructional move thus is an instance in which Pete attends to the disciplinary substance of the group’s thinking. We understand the energy idea that Selena voices as important and fruitful for modeling the scenarios the group considers in the class period; at the same time, Selena’s rule is inaccurate in some situations. That is, energy can flow from objects with less energy to objects with more energy, if the object with less energy has a higher temperature. We note that Selena’s rule was reasonable for the scenarios the group had considered in worksheet up to that point (which involved objects of the same material and similar size), though not broadly correct in every possible scenario (e.g., scenarios involving different masses and materials). It seems likely that Pete noticed this and thought it would be important to address this discrepancy. Pete then responds with a “thought experiment” that incorporates the group’s ideas about energy flow and seems to be chosen to press the students’ thinking further toward a key learning goal: differentiating between energy, temperature and heat transfer. We interpret the “thought experiment” Pete presents here as a key instructional move that takes up and pursues the disciplinary content of the students’ thinking. We note that this question would be very challenging to answer in a canonically accurate and quantitative way. However, it calls back to the scenario discussed at the beginning of the class (Fig. 1), and we suspect that Pete anticipates they have (qualitative) ideas about this question from the previous discussion.

**Sam:** The oil? Oh wait no, it’s the water.

**Pete:** The water has more energy? Okay. Which has more temperature?

**Stephanie:** Oil.

Pete again revoices and affirms Sam’s idea, then checks the group’s thinking about the temperature of the two liquids. Since Pete gave specific temperatures for the oil and water in his question, we interpret this new question as a reminder intended to draw on ideas he knows the students have used, rather than a question to elicit their ideas. This exchange sets up a new question that connects their ideas:

**Pete:** Which way does the energy go?

**Sam:** Would the water be going to the oil? To even out?

**Pete:** So they would even out, according to your rule. Which way would it go?

*There is a pause as the group thinks about the question.*
Stephanie: It would go from oil to water.
Sam: But the oil doesn’t have as much energy as the water.

Pete’s repeated questions about the direction of energy flow, in conjunction with his affirmation of the rule that the temperature will “even out,” takes up and coordinates the students’ ideas that the oil has a higher temperature than the water, that the oil has less energy than the water, and that energy flows. This sequence of questions continues the conversation in a way that brings the group to a “vexation point,” [27] or recognition of a gap in understanding that catalyzes sensemaking. This is evidenced by the pause in conversation before Stephanie responds and by Sam’s comment, “but the oil doesn’t have as much energy…” The conversation continues as Pete affirms what Sam points out:

Pete: that’s the conundrum, right?  
Stephanie: But the temperature is higher, so the temperature has to even out. And that we know, I know that temperature goes from high to low. No matter what.
Selena: So it’s not energy, but temperature that goes from high to low.

Pete: Does the temperature go?
Selena: Or not the temperature, the energy makes it so that the temperature –  
Stephanie: The energy is transferred from like, from a, like a higher temperature to lower temperature. So if you wanna relate those two…
Selena: It moves, it moves, wait … energy moves. Yeah. From an object that is higher to the lower temperature.

Pete: So from hot to cold.
Stephanie: Yeah.

Here Stephanie and Selena reformulate their statement, saying that instead of the energy moving from high to low concentration, the temperature moves towards equilibrium. Pete’s question “does the temperature go?” seems to be chosen to push Selena’s statement toward canonical correctness. Even though this question does not come across as open-ended, we see it as responsive to the ideas about energy flow and temperature change that the students have used throughout their conversation. Ultimately, this question helps Stephanie and Selena to connect their ideas: energy transfers from high to low temperature objects.

Following this interaction, the group concludes that water holds more energy than oil at the same temperature because it has a greater heat capacity. They return to the two-block scenario and clarify that the final temperature is halfway between the initial temperatures only because the blocks are identical. Stephanie says, “In this case, energy and temperature can be related and will be like a change in energy will have the same change in temperature for both.” That Stephanie and Selena articulate a refined rule for heat transfer suggests that the thought experiment was successful in connecting and extending their ideas. We also infer that their conversation with Pete ultimately supported the group to resolve the question that was “bothering” Stephanie earlier in the class period. Thus, Pete’s instructional moves in this conversation (particularly the water and oil thought experiment) leveraged students’ own ideas, advancing them toward an introductory physics learning goal.

V. DISCUSSION/CONCLUSION

In this example of responsive teaching, Pete asked a series of questions that first drew out then leveraged, connected, and refined students’ thinking. Pete’s instructional moves centered around his thought experiment about the energy and temperature of equal amounts of oil and water, building on a similar question he posed at the start of the class. This question successfully targets the students’ thinking – we see that the question draws on ideas they have articulated previously and relied on throughout their discussion. It successfully connects several of these ideas, such that the group refines the rule they had previously articulated in a way that more closely approaches a canonical physics understanding. This process of understanding student ideas and teaching in a way that pursues student thinking is one of the hallmarks of responsive teaching.

This episode suggests that one way instructors may be able to implement responsive teaching practices in university science courses is to plan ahead questions and/or thought experiments, like the one Pete uses here, that can be flexibly deployed to respond to and build on students’ thinking. These planned questions might anticipate common student ideas so that the instructor could respond to these ideas through a responsive teaching framework.

We note that the instructional materials used in the class were designed to provide open-ended scaffolding for student thinking, and Pete drew on that structure (asking for “rules”) as he opened the conversation we analyzed. We hypothesize that the open-endedness of the worksheet and its design to elicit a range of common student ideas were important for this interaction. This structure may have supported Pete to plan for and use questions like the one used in this episode. Future work could explore whether materials that provide open-ended scaffolding, like ACORN Physics worksheets, are particularly supportive of responsive teaching practices.

ACKNOWLEDGEMENTS

This work was supported in part by the National Science Foundation under grants 1914603 and 1914572. We would like to thank Al Snow and Amy Robertson for providing valuable feedback on our analysis and writing.

While Pete’s questions up to this point bear some resemblance to an “elicit-confront-resolve” strategy [28], we do not see any evidence that Pete intentionally chose questions to draw out the incorrect idea that “energy goes from high concentration to low concentration.” Instead, we see Pete responding to the students’ emergent idea with a series of questions that make visible some inconsistencies in their reasoning. The key distinction we see is that Pete’s questions are guided by and build on students’ in-the-moment thinking.
Students’ interpretations of disciplinary convention with the first law of thermodynamics

Alexander P. Parobek
Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, IN, 47907

Marcy H. Towns
Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, IN, 47907

The transfer of knowledge within and across disciplines remains a compelling challenge for modern STEM education and further research is needed to expand on the student-exhibited cognitive and affective gains achieved by innovative cross-disciplinary STEM instructional techniques. This study seeks to support cross-disciplinary STEM instruction and learning by investigating how students use the first law of thermodynamics, a crucial principle to the crosscutting concept of energy and matter, to bridge across disciplinary boundaries. An interview study was undertaken wherein chemistry-, engineering-, and physics-major students addressed a common set of conceptual prompts written with different field-specific conventions. This report focuses on students’ interpretations of the provided forms of the first law and work equations between prompts. Emergent findings demonstrate field-specific interpretations of arbitrary differences in convention and strong barriers to transfer. Derived implications inform suggestions for scaffolding across such disciplinary differences and for future work in this area.
I. INTRODUCTION

Modern education research at the undergraduate level has demonstrated the positive impact that multidisciplinary, interdisciplinary, and transdisciplinary instructional approaches can have on students’ learning and attitude towards learning STEM [1-3]. Supporting progressive teaching models should therefore be a major focus of education research and should be guided by a cross-disciplinary set of frameworks. The Next Generation Science Standards outlines the crosscutting concepts (CCCs) as the common tools and lenses used to bridge across the disciplines of STEM [4,5]. Despite the critical role of the CCCs in defining the topics which bridge the disciplines of STEM, little work has been conducted to date on the CCCs at the undergraduate level [5,6]. This interview study seeks to support such integrated models of STEM education by investigating students’ abilities to apply the first law of thermodynamics, a crucial principle of the CCC of energy and matter, across the fields of science and engineering. In particular, the findings summarized herein focus on students’ conceptualization of differences in disciplinary convention when addressing the first law equation.

Bridging from one disciplinary context to another may be viewed as transfer of learning [7]. A student-centered transfer of learning framework is applied in this study given the notable advancements of such frameworks in modeling and supporting transfer [8]. Recent studies have highlighted the critical role that epistemology plays in governing whether transfer emerges in a productive or unproductive fashion [9,10]. Addressing epistemology in the classroom is particularly challenging given that discipline-specific epistemic viewpoints emerge across traditional course environments and these messages may conflict depending on the context [11,12]. Therefore, building students’ productive epistemic performance should be a major focus of future cross-disciplinary research [13].

Within the physics education research literature, significant advances have been made to understanding the guiding epistemologies that impact students’ applications of mathematical representations. Redish & Gupta [14] introduced a seminal model for physical modelling that highlights four key skills that a scientist must engage in to effectively describe the physical world with mathematics. When engaging in modeling, students have been shown to commonly encounter barriers when applying only a limited set of modeling skills [15,16]. Such barriers are highlighted by students’ stated reasons or “warrants” for adopting certain skills and the analysis of these warrants has revealed distinct guiding epistemologies in how students frame equations in a problem-solving context [17]. Unproductive guiding epistemologies often emerge from perceived authority [18] and the resulting trust in such authority over intuition in certain contexts [19]. However, flexibility in navigating between different epistemic viewpoints is notably desirable and the mark of expert-like modeling behavior, especially in the case where the modeling context is counterintuitive [20].

This study seeks to build on prior transfer and physical modeling research by examining how students leverage disciplinary differences in convention, when addressing the first law of thermodynamics, to bridge across disciplinary boundaries. For the purposes of this study, “disciplinary boundary” is defined as the set of systems [21], language [22], and notation [23] used to frame a problem-solving context. An interview study was conducted to engage students in solving a set of common conceptual first law problems for which the systems, language, and notation were varied across the sample of disciplines studied. Data analysis was focused on identifying how students realized the different disciplinary conventions to know about the first law problem-solving scenario. The guiding research question for the applied analysis was: “How does notational convention impact students’ approaches to solving problems pertaining to the first law of thermodynamics?”

II. METHODS

A. Framework

The Dynamic Transfer (DT) framework [24] served as the methodological and theoretical basis for the applied methods and analysis. As a student-centered transfer of learning framework, DT models the process by which a student is primed by an interviewer to make knowledge available to themselves within an interview setting. A students’ problem-solving expectations, the context they identify, and the ideas they use are all viewed as fine-grain knowledge elements or tools. The distinction of tools within DT is consistent with a manifold ontology of knowledge and the resources framework [25,26]. Where DT differs and expands upon these foundational perspectives is the structure of the model as it pertains to the unique context of the interview setting. The role of the interviewer in “priming” students to adopt a particular epistemology and the process of using the provided context to construct knowledge may all be modelled through this lens. As such, the application of DT in the case of this study may be viewed as an epistemic game [15] whereby the applied methods investigate “how” students access what knowledge they have rather than ascertaining “what” knowledge they have.

B. Interview prompts and protocol

Three discipline-specific interview problems were developed that tasked students with determining the change in internal energy for a piston-cylinder system following described heat and work processes. The developed problems were printed on paper and students were asked to draw a picture of the described system and to solve the problems in a think-aloud style. Each prompt had the same base structure as summarized:
1. Description of the system
2. Draw the system
3. Heat and work process descriptions
4. Determine the internal energy of the system
5. Provided first law and work equations
6. Problem question and MC answer choices

For each prompt, the systems, language, and notation defining the context were varied to incorporate the disciplinary conventions of thermodynamics in chemistry, engineering, and physics instruction. Therefore, each interview prompt may be viewed as variations of the same thermodynamics problem with arbitrary alterations in disciplinary context that provide the “task distance” for this transfer experiment [27]. Prompts were developed by first drawing from relevant textbook materials in each field and then vetting the prompts to align with classroom-specific practices. This report focuses on students’ interpretations of the various disciplinary conventions associated with the first law and work equations across each prompt (see Table I).

Table I. Provided equations for interview prompts.

<table>
<thead>
<tr>
<th>Prompt</th>
<th>First Law Equation</th>
<th>Work Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics</td>
<td>( \Delta E_{\text{int}} = Q - W )</td>
<td>( W = \int_{V_i}^{V_f} p,dV )</td>
</tr>
<tr>
<td>Chemistry</td>
<td>( \Delta E = q + w )</td>
<td>( w = -P\Delta V )</td>
</tr>
<tr>
<td>Engineering</td>
<td>( Q - W = \Delta U + \Delta KE + \Delta PE )</td>
<td>( W = \int p,dV )</td>
</tr>
</tbody>
</table>

All students addressed one in-discipline and one out-of-discipline prompts (order varied) in succession. Afterwards, students engaged in a scaffolded transfer phase. During this phase, both prompts were placed side-by-side and students were asked to compare the similarities and differences between both prompts. The purpose of the scaffolded transfer phase was to support students transfer into an unfamiliar disciplinary context by prompting their attention towards the common application of the first law of thermodynamics in each prompt.

C. Sampling and analysis

A total of \( n = 40 \) students were recruited from introductory thermodynamics courses for majors in chemistry (\( n = 10 \)), engineering (\( n = 20 \)), and physics (\( n = 10 \)). All participants were recruited from the same large institution which offered discipline-specific course sequences in each field of interest. Chemistry and physics students only addressed the physics and chemistry out-of-discipline prompts respectively. Analysis was conducted via a previously reported general inductive approach [28,29]. Validity and reliability were established by iterative interrater coding mediated by theoretical discussion.

The relative lack of notational differences between the engineering and physics prompts was reflected in the minimal activation of ideas associated with these differences. Therefore, the discussion of this report will focus on the portion of engineering students who addressed the chemistry prompt (ENG-C) and the remaining interview sample (\( n = 30 \) total). Participant quotations have been designated by an alphanumeric code denoting their discipline and interview number (C1, P3, etc.) and the stage of the interview (chemistry as first prompt: C1, scaffolded transfer phase: ST).

III. RESULTS/FINDINGS

A. Interpretations of first law equation

Disciplinary inconsistencies in the capitalization of heat and work symbols in the first law equation was the most frequently noted attribute that distinguished the chemistry prompt from the physics and engineering prompts (17 of 30 interviews). Most notably, engineering students, such as E7 below, indicated that the lower case heat and work terms of the chemistry prompt suggested that those variables were being considered on an extensive basis:

E7.C1: “I actually, actually it might be the fact that \( Q \) and \( W \) are actually lower case because usually in class you see them with that means is the, there would be the heat transfer and the work done per unit mass.”

E7’s conjecture was common among engineering students and reflects the distinct use of intensive and extensive notation via letter case in the engineering course studied when compared to the chemistry and physics courses of interest.

Conversely, chemistry and physics students expressed confusion when encountering unfamiliar letter case. Both classroom environments implemented a different letter case convention when describing heat and work. However, capitalization rarely altered students’ work on the provided problems as exemplified in P4’s analysis of the unfamiliar chemistry notation:

P4.C2: “Um, it's kind of unusual that \( Q \) and \( W \) are lower case, kind of weirds me out a little bit, cause I'm so used to seeing them as capital \( Q \) and capital \( W \), but I'm just, I'm kind of also assuming that they're just the same variables.”

Here P4 recognizes an irregularity in the letter case used for heat and work, but P4 eventually treats heat and work in the same way as they would if it were upper case. This recognition and approach was mirrored by chemistry students who had similar reactions to the upper case heat and work terms of the physics prompt. Ultimately, the distinguishing factor between students that did or did not associate meaning with the letter case of heat and work was reflective of whether their classroom environments communicated intended physical information through the purposeful use of letter case.

Students’ problem-solving appeared to be more sensitive to their perception of what dependent variable was used to
describe the expression. While each prompt was defined in terms of internal energy, some chemistry students (2 of 10 interviews) read the energy-internal term of the physics prompt to encode for some unfamiliar term that was distinct from the term provided in the chemistry prompt. C3 demonstrates this point below showcasing how unfamiliarity with the energy-internal term and the work equation impacted their interpretation of the equations in the context of the problem:

C3.ST: “Well in Problem 1 the E has I-N-T after it like in the subscript, which means internal normally. And then in the second one it just has Delta-E with no internal, and that equation I know does not deal with just internal. And I haven't seen Problem 1's equation before, at least for the W, which makes me think it's probably only for internal energy.”

Interestingly, C3 appears to be confident that the energy term of the physics problem (Problem 1) signifies the internal energy when reading out the subscript and uses this to infer that the more familiar chemistry term does not signify “just” internal energy (Problem 2). C3’s perception of the dependent first law variables is critical to consider given that 8 out of 10 chemistry students ultimately came to an unproductive assessment of the physics prompt. A previous report has summarized chemistry students’ tendencies to rely on causal-mechanistic reasoning when approaching the physics prompt [28]. Similar instances of uncertainty in declaring distinctions between the dependent variables of the provided first law equations were absent from physics and engineering students’ reflections.

B. Interpretations of work equation

Students across the disciplines reflected on the different forms of the work equation employed in the chemistry problem when compared to the engineering and physics problems (20 of 30 interviews.) As suggested by C3 in the prior section, the general form of the boundary work equation was unfamiliar and tended to lead chemistry students to differentiating the corresponding first law equations across prompts. Conversely, engineering and physics students demonstrated expanded mathematical aptitude when addressing these different forms of the work expression. Engineering students, such as E8 below, commonly pointed out (4 of 10 interviews) that the connection between the two forms of the expression lie in a constant pressure assumption that would allow the pressure to be pulled “out of the integral:”

E8.C2: “[...] because, um, when we're solving problems, we always have to write basic equations and every time for work, the basic equation is integral of P times D-V. And if, if we ever want to make the equation just P-Delta-V we have to be able to pull the pressure value out of the integral.”

E8 discusses the connections between the simplified and general work expression without inferring any new information about the problem. Instead, E8 reflects on their prior experiences in approaching thermodynamics problems and how they always start from the more general form of the expression.

Physics students contrasted from engineering students in their notable tendency to infer attributes of the provided problem and described processes based on the provided work equation (3 of 10 interviews). Consider P3’s comment below when comparing the chemistry and physics prompt equations:

P3.P2: “Again, pay attention to the fact that this work is given as an integral and not just P-Delta-V because it implies that P changes, cause of P does change. P-Delta-V wouldn't work, that's why they changed the form, that's why they changed the equation.”

Unlike E8’s discussion, P3 suggests that pressure is implied to change within the physics prompt given the more general form of the expression provided. This distinction between E8 and P3 signifies a difference in the student-realized meaning of the provided prompts when comparing the students across disciplines.

C. Problem-dependence of equations

During the scaffolded transfer phase, students that used the provided equations to evaluate the problems were asked whether they felt the equations were only relevant to the provided problem or if they could be applied to both problems. Students’ responses were binned into two mutually exclusive coding definitions included in Table II to distinguish whether students saw the equations as dependent or independent to the problem-solving context. Frequencies of each code across the interview sample is included for reference. The sum of these frequencies does not reach n = 10 for each discipline given that not all students utilized the first law equation to solve the provided interview problems.

The “problem as equation dependent” code was observed across the engineering and chemistry sample and was absent from the physics student sample. Most notably, student interviews for which the problem as equation dependent code emerged encountered unproductive barriers within the scaffolded transfer phase (5 of 6 interviews). Each case was marked by an unwillingness to productively apply the more familiar first law and work expressions to solve the out-of-discipline prompt. Only a small portion of students indicating problem as equation independent encountered similar barriers during the scaffolded transfer phase and all were chemistry students (3 of 20 interviews). Chemistry students, during the scaffolded transfer phase, would often cite a lack of familiarity with the equations provided in the physics prompt and would then indicate previously outlined features of the equations as reasons for this uncertainty:
problems: equations after reconciling the differences between both the problem with the more familiar first law and work equations, only one chemistry student (C7) came to evaluate students which avoided using the provided physics familiar first law and work equations. Of the eight chemistry expressions impacts their perceived relevance of the more inherent uncertainty with the provided first law and work equations. These expressions such as deducing that an integral work seed different student-realized interpretations. In particular, engineering, and physics prompts of this study are shown to impact students from the most prevalent unproductive approaches to evaluating the provided problems. Of the eight chemistry students that were unproductive when addressing the physics prompt, only one chemistry student shifted to productively applying the equations provided on the chemistry prompt to the physics prompt. The shortcomings of this stage may be understood when considering the critical role of epistemic agency [30] in governing to what degree students are able to build knowledge in a learning space. A student may encounter a barrier when evaluating a problem out of discipline if they conclude that signs of ambiguity or unfamiliarity are the result of a personal lack of understanding. While metacognition on what one has learned and needs to learn is useful [31], the arbitrary variation of systems, language, and notation in the case of this study provides evidence for the emergence of an epistemic barrier derived from perceived authority.

These findings further support the previously outlined call to vary instruction of the first law of thermodynamics across disciplinary environments to emphasize the conceptual, mapping, and arithmetical power of this fundamental energy and matter principle [28,29]. Furthermore, this report suggests that building productive epistemic performance [9,13] with CCCs may require a general shift towards preparing students both to conceptually grapple with cross-disciplinary topics and to recognize the capacity of physical mathematical relationships, which serve as guiding principles to CCCs, to model reality [17,24]. Future work is needed to better understand the ways in which disciplinary acculturation has impacted students’ abilities to leverage CCCs for the purposes of transfer.

Findings derived from this study are non-generalizable beyond the unique classroom environments that were investigated. The application of the Dynamic Transfer framework in this study restricts the findings to exploring how students realize the provided disciplinary context and does not track how ideas that students activate in these contexts became incorporated into long-term memory.

IV. CONCLUSIONS

The arbitrary alterations in notation across the chemistry, engineering, and physics prompts of this study are shown to seed different student-realized interpretations. In particular, trends were identified in what interpretations students across these disciplines make when encountering an unfamiliar context. Chemistry students were notably keen to express uncertainty when addressing unfamiliar forms of the first law and work expressions and to refrain from applying equations they felt more familiar with. Engineering and physics students, while more able to interpret the provided differences, sometimes associated additional ideas with these expressions such as deducing that an integral work expression was provided to signify a changing pressure.

Most importantly, the inclusion of the scaffolded transfer phase in this study did not appear to significantly sway students from the most prevalent unproductive approaches to evaluating the provided problems. Of the eight chemistry students that were unproductive when addressing the physics prompt, only one chemistry student shifted to productively applying the equations provided on the chemistry prompt to the physics prompt. The shortcomings of this stage may be understood when considering the critical role of epistemic agency [30] in governing to what degree students are able to build knowledge in a learning space. A student may encounter a barrier when evaluating a problem out of discipline if they conclude that signs of ambiguity or unfamiliarity are the result of a personal lack of understanding. While metacognition on what one has learned and needs to learn is useful [31], the arbitrary variation of systems, language, and notation in the case of this study provides evidence for the emergence of an epistemic barrier derived from perceived authority.

The quotation by C9 above demonstrates how their inherent uncertainty with the provided first law and work expressions impacts their perceived relevance of the more familiar first law and work equations. Of the eight chemistry students which avoided using the provided physics equations, only one chemistry student (C7) came to evaluate the problem with the more familiar first law and work equations after reconciling the differences between both problems:

C9.ST: “Um, well Equation #1 or 2 in Problem #1 is more difficult to solve than the problem, than Equation #2 in Problem #2 and again, and the, um, the variables are a different capitalization so they might not even mean the same thing.”

C7.ST: “Like I was taking the, um, idea that from Problem 2, I was taking Equation 2, and from that I was saying that work was positive and then I was taking it and applying it to Equation 1 in Problem 1, which you can't do, you can't mix and match like that because work is found in different ways in both of the columns.”

C7 ultimately comes to a productive assessment of the physics prompt due to similarities they read out between the two prompts during the scaffolded transfer phase. Achieving this outcome notably required C7 to recognize how their ideas about work relate to the first law equations provided in the chemistry and physics prompts.

IV. CONCLUSIONS

The arbitrary alterations in notation across the chemistry, engineering, and physics prompts of this study are shown to seed different student-realized interpretations. In particular, trends were identified in what interpretations students across these disciplines make when encountering an unfamiliar context. Chemistry students were notably keen to express uncertainty when addressing unfamiliar forms of the first law and work expressions and to refrain from applying equations they felt more familiar with. Engineering and physics students, while more able to interpret the provided differences, sometimes associated additional ideas with these expressions such as deducing that an integral work expression was provided to signify a changing pressure.

Most importantly, the inclusion of the scaffolded transfer phase in this study did not appear to significantly sway students from the most prevalent unproductive approaches to evaluating the provided problems. Of the eight chemistry students that were unproductive when addressing the physics prompt, only one chemistry student shifted to productively applying the equations provided on the chemistry prompt to the physics prompt. The shortcomings of this stage may be understood when considering the critical role of epistemic agency [30] in governing to what degree students are able to build knowledge in a learning space. A student may encounter a barrier when evaluating a problem out of discipline if they conclude that signs of ambiguity or unfamiliarity are the result of a personal lack of understanding. While metacognition on what one has learned and needs to learn is useful [31], the arbitrary variation of systems, language, and notation in the case of this study provides evidence for the emergence of an epistemic barrier derived from perceived authority.

The quotation by C9 above demonstrates how their inherent uncertainty with the provided first law and work expressions impacts their perceived relevance of the more familiar first law and work equations. Of the eight chemistry students which avoided using the provided physics equations, only one chemistry student (C7) came to evaluate the problem with the more familiar first law and work equations after reconciling the differences between both problems:

C9.ST: “Um, well Equation #1 or 2 in Problem #1 is more difficult to solve than the problem, than Equation #2 in Problem #2 and again, and the, um, the variables are a different capitalization so they might not even mean the same thing.”

The quotation by C9 above demonstrates how their inherent uncertainty with the provided first law and work expressions impacts their perceived relevance of the more familiar first law and work equations. Of the eight chemistry students which avoided using the provided physics equations, only one chemistry student (C7) came to evaluate the problem with the more familiar first law and work equations after reconciling the differences between both problems:

C7.ST: “Like I was taking the, um, idea that from Problem 2, I was taking Equation 2, and from that I was saying that work was positive and then I was taking it and applying it to Equation 1 in Problem 1, which you can't do, you can't mix and match like that because work is found in different ways in both of the columns.”

C7 ultimately comes to a productive assessment of the physics prompt due to similarities they read out between the two prompts during the scaffolded transfer phase. Achieving this outcome notably required C7 to recognize how their ideas about work relate to the first law equations provided in the chemistry and physics prompts.

IV. CONCLUSIONS

The arbitrary alterations in notation across the chemistry, engineering, and physics prompts of this study are shown to seed different student-realized interpretations. In particular, trends were identified in what interpretations students across these disciplines make when encountering an unfamiliar context. Chemistry students were notably keen to express uncertainty when addressing unfamiliar forms of the first law and work expressions and to refrain from applying equations they felt more familiar with. Engineering and physics students, while more able to interpret the provided differences, sometimes associated additional ideas with these expressions such as deducing that an integral work expression was provided to signify a changing pressure.

Most importantly, the inclusion of the scaffolded transfer phase in this study did not appear to significantly sway students from the most prevalent unproductive approaches to evaluating the provided problems. Of the eight chemistry students that were unproductive when addressing the physics prompt, only one chemistry student shifted to productively applying the equations provided on the chemistry prompt to the physics prompt. The shortcomings of this stage may be understood when considering the critical role of epistemic agency [30] in governing to what degree students are able to build knowledge in a learning space. A student may encounter a barrier when evaluating a problem out of discipline if they conclude that signs of ambiguity or unfamiliarity are the result of a personal lack of understanding. While metacognition on what one has learned and needs to learn is useful [31], the arbitrary variation of systems, language, and notation in the case of this study provides evidence for the emergence of an epistemic barrier derived from perceived authority.

These findings further support the previously outlined call to vary instruction of the first law of thermodynamics across disciplinary environments to emphasize the conceptual, mapping, and arithmetical power of this fundamental energy and matter principle [28,29]. Furthermore, this report suggests that building productive epistemic performance [9,13] with CCCs may require a general shift towards preparing students both to conceptually grapple with cross-disciplinary topics and to recognize the capacity of physical mathematical relationships, which serve as guiding principles to CCCs, to model reality [17,24]. Future work is needed to better understand the ways in which disciplinary acculturation has impacted students’ abilities to leverage CCCs for the purposes of transfer.

Findings derived from this study are non-generalizable beyond the unique classroom environments that were investigated. The application of the Dynamic Transfer framework in this study restricts the findings to exploring how students realize the provided disciplinary context and does not track how ideas that students activate in these contexts became incorporated into long-term memory.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under DUE Grant #: 1948981.

Table II. Problem-dependence of equation codes and frequency.

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>Count by Discipline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem as equation dependent</td>
<td>Statement that the provided equations in a prompt is specific to that prompt when comparing the equations provided in both prompts.</td>
<td>Chem: 3, Phys: 0, Eng-C: 3</td>
</tr>
<tr>
<td>Problem as equation independent</td>
<td>Statement that the provided equations in a prompt can be applied in either prompt when comparing the equations provided in both prompts.</td>
<td>Chem: 7, Phys: 7, Eng-C: 6</td>
</tr>
</tbody>
</table>
LA program as a driving force for identity development through access to ideational resources

Xandria R. Quichocho (they/them)+, Simone Hyater-Adams (she/her)*, Eleanor W. Close (she/her)+

Physics Department, Texas State University, 601 University Dr San Marcos, TX, 78666
*MEGA Imagination

In our research we invite multiply-marginalized physics students—Black, Indigenous, Women of Color and LGBTQ+ women—at Hispanic Serving Institutions to participate in semi-structured interviews about their experiences in physics. Interviewees were asked about their perceptions of their physics environments, available support systems, and how they would describe a physicist. Participants from Texas State University described multiple positive impacts from the Physics Learning Assistant (LA) Program. We analyze these narratives using the Critical Physics Identity framework to see in what ways LA participation overlaps with participants’ racialized and/or queer experiences studying physics. We present a case study of a single participant, focused on the ways in which the LA Program provided her access to relational resources (aspects of relationships that impact one’s connection to physics) and ideational resources (aspects of an idea that impact one’s connection to physics), and how these resources supported her identity negotiation and positive physics identity development.
I. INTRODUCTION

The goal of this paper is to begin to understand the ways the Texas State University (TXST) Physics Learning Assistant (LA) Program affects the physics identity development of multiply marginalized students, by analyzing interview data using the lens of the Critical Physics Identity framework (CPI) [1,2]. In this paper we present a case study of Frankie, a Hispanic woman interviewed shortly after she graduated from TXST with a B.S. in Physics, focusing on her experiences in the introductory course sequence for physics majors. We will illustrate how her relationships with peers, LAs, and instructors provided her with relational and ideational resources that shifted her physics practices and supported her positive physics identity development.

II. BACKGROUND & METHODS

The case study presented here draws from an interview of a TXST student. As discussed in prior work [3-5], TXST is a Hispanic Serving Institution where the racial demographics of the university are mirrored in that of the physics department; and where nearly half of declared physics majors identify as a racial or ethnic minority, allowing for students to study in a diverse learning environment [4].

This analysis is part of a larger interview study designed around the question “How do multiply minoritized/multiply marginalized physics students perform and discuss the intersections of their physics identities with their social identities?” [3-5]. The interviews in this study were video recorded and conducted by the first author, who, in being a racial/ethnic minority and a queer person, shares many of the same minoritized and marginalized identities. Interviews focused on the interviewees’ narratives about their experiences studying physics, and how their learning environments impacted their physics identity.

The team iteratively coded the transcripts using the Critical Physics Identity (CPI) framework [1,2] to explore the relationships between Physics LA Program structures and practices, interviewees’ interactions with the Physics LA Program, and their physics identity development.

A. TXST Physics LA Program

The TXST Physics LA Program is based on the model developed at University of Colorado Boulder [6], with an emphasis on building and supporting a sense of community [7]. Unlike the opt-in model used at CU-Boulder in which individual faculty choose whether to apply for LA support, at TXST LAs support all sections of the three-course introductory physics sequence for majors. The primary LA role is assisting small groups of students during the “lecture” class meetings, often while groups are working on materials from Tutorials in Introductory Physics [8]. All LAs and LA-supported faculty meet together for collaborative weekly preparation sessions on Friday afternoons. The first ~30 minutes of this time is spent with LAs and faculty from all courses discussing common issues and concerns; this is followed by time for each course team to work together to debrief and prepare for the following week. At the time of these interviews, physics was the only TXST department with LAs, and the required pedagogy course for first semester LAs was a physics-specific course, taught by the third author.

While all majors are welcome to apply to the Physics LA Program, typically more than half of applicants are physics majors or minors. Because the introductory sequence for physics majors also serves majors from other, larger departments (e.g., engineering, chemistry), most physics majors have the opportunity to serve as LAs, and in fact the majority of our B.S. graduates have LA experience by the time they complete their degrees. We also make a special effort to recruit women physics majors into the LA program. Previous research has found that participation in the LA program supports students’ development of physics identity [7], and that being a part of the program helps them find “success together” [3].

B. Critical Physics Identity Framework

The Critical Physics Identity (CPI) framework [1,2] was developed by the second author to deconstruct the experiences of Black physicists to show how structures and systems in the field impact their physics identity development. The CPI framework combines two identity frameworks from the education research field: Hazari et al.’s physics identity constructs (interest, performance, competence, and recognition) [9] originating from Carlone & Johnson’s science identity framework used to indicate the “stuff” of a physics identity [10], and Nasir’s racialized identity resources (material, relational, and ideational) used to examine how resources provided by the localized environment can encourage or discourage Black students’ connections to a practice [11]. With the CPI, we can use these combined constructs to examine more deeply the complex factors that contribute to how marginalized physicists view themselves in the context of, and connection to, the physics field.

The CPI weaves together the Nasir and Hazari frameworks to more accurately understand the lived experiences of racialized physicists. Under the umbrella of the CPI framework, the identity resources and physics identity constructs work with each other to describe a person’s physics identity. While this framework was originally operationalized using the narratives of Black physicists, we find the CPI an extremely useful lens for critically examining the many institutional, systemic, and cultural ways physics perpetuates racism, sexism, homophobia, and transphobia [2].

By analyzing the sections of the interviews that discuss the TXST LA Program using the CPI, we are able to discern
what components of the program are supporting multiply marginalized populations in their physics identity development. Here, we use the framework to gain a deeper understanding of Frankie’s experiences and their impacts as she describes her journey through her undergraduate physics courses.

In Frankie’s narrative, we focus on the relational and ideational resources that Frankie uses to negotiate her place in, and connection to, physics and more specifically the physics classroom. *Relational resources* are defined as aspects of a relationship with others in the physics context that impact one’s connection to physics. A single relationship has the capacity to provide an individual with multiple relational resources in different ways and in different times of their physics journey. Similarly, *ideational resources* are defined as aspects of an idea (e.g., ideas about oneself, one's place in physics, or what is valued in physics) that impact one’s connection to physics. Ideational resources are not solely developed in the physics environment, and students often come into the classroom with a number of ideational resources already [1,2]. Often these ideational resources inhibit students’ connection to physics; we often see this dissonance in how students describe a physicist, or in how they describe their own physics identity.

We code elements of Frankie’s story using these resources to illuminate the ways in which both the introductory course structures and the relationships she develops in these classes impact how she positions herself in the classroom, her understanding of what is valued, and how these ideas shift over time.

### III. ANALYSIS

Frankie, who self-identified as a Latina/white and bisexual woman, was interviewed shortly after graduating from TXST with a B.S. in physics. During her time as an undergraduate at TXST, Frankie completed the LA-supported calculus-based introductory physics sequence and served as an LA for multiple semesters. In her interview, Frankie shared a narrative about her time in the introductory course sequence that gives us rich insight into how the resources to which she had access helped her increase her connection to the field and develop her physics identity as a student and as an LA.

Frankie’s description of her first three semesters in the TXST physics department is a story of increasing engagement: with peers; with student-centered structures in courses; with physics content; and with the LA program. Because the TXST Physics LA Program supports all sections of all three courses in the introductory sequence for majors, Frankie had multiple semesters of support for developing positive relationships with peers, and continuous reinforcement of new ideational resources about how to engage with physics and what it means to be a good physics student.

Frankie arrives in her first physics class with specific ideas about what characteristics are necessary to be a successful student, and specific perceptions of who physicists are. She knows that women are highly underrepresented in physics, and recognizes that she is one of very few in the classroom. It is important to note that Frankie speaks less about her race and ethnicity than her gender in our interview, possibly because of the department’s large representation of Latina/o students; Frankie was more likely to be the only woman in a class than the only Latina/o student.

As her comfort increases as she understands the expectations of the space, her ideas of what it means to engage in the classroom and be a successful student shift. She begins to intentionally push herself to be more active in this space, and grows more confident and more outspoken in her peer groups. This extends into becoming an LA herself, and making herself do things as a student that she knows she needs to get better at to be a good LA. Joining the LA Program amplifies the ideational resources she first encounters as a student. By the end of her introductory sequence, Frankie’s new ideational resources have driven her to radically shift the way she practices physics and the way she connects to physics, reflecting a strengthened physics identity. In the following sections, we deconstruct Frankie’s story to go into more detail about the factors that supported Frankie’s development over the introductory sequence.

#### A. “Kinda shy at first”

Frankie describes herself as “shy at first,” coming into her first introductory physics course with the intention of “sitting in the front and being that good student.” When she realizes that the in-class work will be in small groups with other students, she immediately seeks out a partner who shares some of her identities as a Latina/white woman, who she calls, “the little Hispanic chick.” Frankie took up little vocal space in discussions, rarely taking the initiative to speak first and content to follow the lead of the more outspoken group members. Still, because of her performance in class, her professor recognized her competence and encouraged her to apply to the LA program; however, Frankie did not feel confident enough to do so at the time.

Frankie’s initial ideational resources related to her own positioning in the classroom, her personal characteristics, and her perceptions of what practices are valued in the physics classroom inform how she plans to engage in her first introductory class. She immediately recognizes that her identity as a woman puts her in the minority in her majority-male class: “as a woman, in my physics classes obviously there wasn’t a whole lot of us, so I was kinda shy at first.” Her initial shyness aligns well with what she believes are the characteristics of a good student, being quiet and attentive. She did not enter the class with an expectation that she would have to engage with her peers in a sustained way. However,
as the TXST LA Program is integrated fully into every introductory course, students are asked from the start to form groups and talk to each other immediately challenging Frankie’s initial ideational resources:

“I’m kinda shy in general, I’m not one to go up to people and start a conversation. I remember just sitting in the front and being that good student, being like I’m just here for class. But then in physics, you have to work with people, so they’re kind of like, ‘Hey, you have to talk to somebody.’”

In this quote we see the start of a shift in how Frankie sees her connection to physics in the physics classroom, primarily with the idea of what practices are valued. In a more traditional lecture setting Frankie would be completely right about what is valued, but she realizes that in this space the way to be a “good student” is talking and working with people to make sense of physics concepts. This is somewhat at odds with how Frankie sees herself, which we coded with the ideational resource personal characteristics: Frankie sees herself as someone who is unlikely to initiate conversation.

Frankie negotiates this dissonance by seeking out specific individuals with whom to work:

“In our group I at least always had one other lady with me. I think that was always my key, I was like, ‘Okay, if I have to work with a group I’m gonna try to work with either someone I know,’ but I was like, ‘Oh, I don’t know anybody,’ so then I was like, ‘Okay, I’m gonna try to work with a lady.’”

This is a key element in understanding the relationships that Frankie seeks, which provided the relational resources that give her access to additional ideational resources during her introductory course journey. One of the first relationships we see is that between Frankie and someone she describes as “the little Hispanic chick who’s also pretty fair skinned like my…” This relationship persists throughout her first semester Mechanics course and into her second semester E&M course. In talking about her second semester in the intro sequence, Frankie describes how this relationship supported her:

“We happened to be in the same E&M class the following semester, so that was really great. Then I was like, ‘Okay, we’re comfortable, we can work together.’ …I could see the difference…after one semester, I guess I started feeling more comfortable with myself and in the classroom, because I was like, ‘Okay, we have to do group work. We’re all here doing the same thing,’ I got over it. I don’t know if it was because I also had someone in my group that I knew, that I knew would back me up…but I felt like I could speak out more.”

Her relationship with this student who had similar racial and gender identities as her impacted Frankie’s connection to physics, providing a relational resource that increased Frankie’s willingness to engage in group interactions since she had someone she “knew would back [her] up.” We see that this relational resource helped Frankie situate herself in the practice of doing group work with others, and led to her speaking up more, enacting practices consistent with her new ideational resource of what is valued in physics. Later in the interview, Frankie reiterates the importance of supportive peer relationships for engaging in meaningful learning during group work.

B. “This is what we should be doing.”

Frankie also describes her relationship with another member of the Mechanics course, the “EE (Electrical Engineering) major.” This EE major “very much took charge” of the group conversations, which is a personal characteristic Frankie says she did not have; rather, she was “glad to follow along” in the group, having “more times where I was just like, ‘Oh, okay, that sounds like that’s right.’” At this point we see Frankie leaning into the personal characteristic resources she came into the class with, which are more aligned with where she is comfortable, while working on negotiating her position and place in the group.

From her actions in this relationship with the EE major in Mechanics, we are able to see a clear shift when she starts to discuss her position and characteristics in later semesters. In Frankie’s second semester of the introductory sequence, Electricity & Magnetism (E&M), she is able to continue her partnership with the “little Hispanic chick” from Mechanics. In addition, she has a [Latina] woman as the course instructor, which also makes her feel comfortable in the classroom. Here she finds herself speaking up more in class, feeling more able to take up space in group discussions and advocate for different ways of thinking. There is a slight change in how Frankie interacts with her group in E&M. Where in Mechanics she primarily played support, agreeing with the other members of the group, in E&M she feels able to speak up more, taking up more verbal space:

“The following semester [after Mechanics] I definitely felt like I was like, ‘No, I think this is how we do it, this is what we should be doing.’”

We can see how her internal personal characteristics have become more in tune with what she sees is valued in this environment: engaging in group discussions, speaking up, offering new ways to approach problems. Along with this, she is beginning to reposition herself within her physics classes as someone who not only learns, but teaches.

Acting on her increased confidence in “how we do it” and on encouragement from one of her LAs in E&M, Frankie applies to be an LA for the following semester. She is accepted and works as an LA while also enrolling in her final introductory course, Waves & Heat.

At the same time, Frankie is aware that she still has room to grow in aligning herself to what she sees is valued in the physics department. There is a conscious action taking place—she has models of the characteristics she wants to have in the LAs that work in the classes she is in, and she applies to be an LA knowing that she “has to apply [herself] more, [she] has to be more comfortable talking.”
C. “For the children.”

In Waves & Heat, Frankie has none of the peers she had worked with in the previous two introductory courses. The “little Hispanic chick” Frankie has in the past two semesters of the intro sequence isn’t in her Waves & Heat class, so Frankie tries to find other women to work with. Sitting in the back of the Waves & Heat classroom is a group of all women that includes two women she calls her “physics moms.” Frankie wants to join them, but the “girl group” is full, and they sit in the back of the classroom where she prefers the front, so she looks for new folks to work with.

She finds Nathan (pseudonym), who is a transfer student, and who, in her words, is “a big dude,” and “intimidating, because he’s just very serious.” We note here that Nathan is also Hispanic, but that Frankie does not mention this in her interview as a factor in choosing to work with him. Like Frankie, Nathan doesn’t know anyone else in the class, he is new to the university as a whole. Frankie is excited because, “[He’s] not gonna be able to leave me to go talk to [his] friends.” Nathan’s newness helps Frankie feel comfortable getting to know him and working with him:

“Because he was new, we had a lot to talk about, so it made me... I felt like I was able to get comfortable with him enough to where we always worked together. I could be like, ‘Hey, I don't know what the fuck I'm doing, come help me.’ where with some people you don't really know, you're not comfortable being like, ‘Hey, I don't know what the fuck I'm doing, show me.’ ...But [Nathan’s] really sweet, and he really knows his stuff, so he was able.”

This is Frankie’s third semester in an LA supported course, but Nathan’s first. Although she’s lost the relational resource provided by her comfortable relationship with the woman she worked with throughout Mechanics and E&M, the collective value of group work itself is an ideational resource strong enough for Frankie to push herself to invest in a new relationship until it becomes similarly comfortable.

Throughout her time in the introductory sequence, we see Frankie’s incoming ideational resources shifting and changing into ones that align more closely with the values of the LA Program. Her interactions with the program as a student and, eventually, as an LA, act as a motivator for her to continue developing these skills until she feels confident enough to do it for others:

“It was also me trying to push myself. I was like, ‘Okay, I have to be an LA, I have to talk to people I don’t know all the time, I have to be able to work in my own group doing all of these things before I can do it for the children.’”

Being an LA while still in the introductory course sequence has her wanting to work on her confidence in speaking up, both in her own class and in the class for which she is LAing. Her motivation to improve so she can “do it for the children” implies that she thinks it is important to work on speaking up not just in order to be a successful student, but also as a valuable thing to do with the students she is teaching. The resources Frankie draws upon that support parts of her physics identity internally she is then able to provide in order to nurture and support beginning physicists, helping them engage in the same meaningful collaborative learning environment she experienced.

Frankie’s participation in the LA Program reflects a growing connection to physics through a commitment to the idea of student-centered community and collaborative physics learning, seen as she aligns herself with the ideational resources provided by the TXST LA-Supported physics classroom. We interpret this as a positive growth in her physics identity.

IV. DISCUSSION AND TAKEAWAYS

Because Frankie, like every TXST physics student, experienced LA-supported introductory courses, she gained access to relational and ideational resources that supported the growth of her physics identity. Frankie leveraged these resources in a way that helped her shift her ideas about how she wanted to engage in physics with her peers, ultimately deciding to become an LA herself.

Frequent opportunities to interact with people at different levels of power is one way that the TXST LA Program creates an environment where students can be vulnerable enough to get out of their comfort zone. Frankie was able to build the relationships she talked about in her interview because of the LA classroom model of collaborative group learning. In doing so she was able to evolve her ideas about who she is and how she does physics in and outside of the classroom to ultimately develop a stronger physics identity and a stronger connection to the TXST physics department.

Frankie’s story teaches us about the impact that a shift in ideational resources can have on the development of one’s physics identity, and the influence that relationships can have on the access one has to valuable ideational resources. This is important for all physics students, but particularly critical for those who are multiply marginalized in the field, whose ideational resources often look different than those of their cisgender, heterosexual, white male peers from the larger story that is popularly told about who does physics. These negative ideational resources may prevent stronger connection to physics both prior to beginning their studies, and during their physics journeys.

From this analysis we are able to see the ways a comprehensive LA program is an environment that can provide valuable relational and ideational resources to help multiply marginalized students strengthen their positive physics identity and thrive in the physics community.

ACKNOWLEDGEMENTS

This work has been supported in part by NSF grants 1557405 and 1928596.


Analogies are known to be powerful tools for making sense of unfamiliar ideas in terms of already understood concepts. Science students regularly encounter unfamiliar ideas, such as microscopic objects that are invisible to our everyday experience and behaviors dictated by quantum mechanics. An understanding of basic concepts of quantum mechanics is useful in many disciplines, especially with the growing field of quantum information sciences and technologies. Physics researchers often use analogies in their own research and science communicators use them to make quantum ideas accessible to K-12 students and across STEM disciplines, but analogy use in upper-division teaching has been less researched. Our research goal is to understand how analogies are used to teach quantum mechanics, and specifically, what prior knowledge is used as a basis for analogies within two widely used quantum mechanics textbooks. This textbook analysis shows the most common bases for analogies include: mathematical structures from linear algebra, which are applied to model quantum systems; everyday life examples, which are used to make quantum systems more familiar and understandable; and macroscopic classical phenomena, which are used to highlight differences between classical and quantum mechanics. We also find authors use different conventions, based on the various cue words that authors use to indicate analogy-based reasoning. In the STEM classroom, this research has implications for enhancing student learning about abstract topics in science.
I. INTRODUCTION

Finding ways to enhance education in physics, particularly in quantum mechanics, is important in light of the growing multidisciplinary field of quantum information science [1–3]. As a result, efforts are being made to make quantum ideas more accessible to the general public, K-12 students, and across STEM disciplines [4–7]. In this paper, we investigate the ways that analogies may be used to enhance teaching of quantum physics. Analogies come in a variety of different forms and can be used in many contexts. In physics education, analogies are widely used as a way to make the unfamiliar world of physics more understandable. Indeed, sophisticated scientific ideas are often explained in pop science books in a way that makes them more accessible and understandable to the general public.

Quantum mechanics is viewed as a more advanced physics subject, traditionally taught after a student has completed several physics classes, such as classical mechanics and electricity and magnetism. This is potentially because students can benefit from learning these subjects beforehand and use that prior knowledge as a learning resource. Still, quantum mechanics can be unintuitive to students, since the rules that govern quantum systems (e.g., atoms, quantum bits) are very different from the rules that govern objects in everyday life and vary from the rules from previous physics classes. Hence, analogies are especially suited to teach quantum mechanics because they are specifically made to bridge prior knowledge, intuition, and experiences to new knowledge. As such, this study analyzes analogy use in two common quantum mechanics textbooks. The following are the research questions we want to address with this analysis:

• What, if any, are common words and phrases authors use that indicate the presence of an analogy?
• What bases of analogies are commonly used in quantum mechanics textbooks?

II. BACKGROUND

Students learn in a variety of different ways, including but not limited to analogies. Indeed, the authors of ABCs of How We Learn note that “Analogies help people learn principles and apply those principles in new situations” through recognizing a common underlying structure, despite surface differences [8]. Thus, analogies are tools for learning through facilitating conceptual understanding, and several specific methods have been formed to help create and understand analogies [9–11]. Further, analogies are generally agreed to be mappings from a base of familiar knowledge to a target of unfamiliar knowledge [9, 10, 12]; for the purpose of this analysis, this is the definition of analogy we use. The usefulness of analogies is demonstrated through the following research, which cover a wide range of scientific contexts and usages.

In an analysis of analogies in physics textbooks, Köhrhasan and Hdır indicated that analogies are “suitable for teaching scientific concepts by comparing an unknown with a known” [13]. Similarly, Podolefsky and Finkelstein demonstrated that analogical scaffolding substantially increases student comprehension and corresponding test scores in upper-division electricity and magnetism. Overall, student learning was seen to increase with analogy use compared to lecture-style teaching without analogy use [11]. Further, Clement investigated how analogies may be used in introductory mechanics as a tool to gauge the level of current student knowledge and understanding and ease the transition into new content [14].

Specifically related to quantum mechanics, Wittmann and Morgan focused on the integration of analogies within lecture, emphasizing student experiences in a non-majors quantum physics course. Through emphasis of everyday experiences and situations, analogies aided student sense-making about quantum mechanics [15]. Schermerhorn et al. investigated the use of analogy-based tutorials to teach students upper-division quantum mechanics based on students’ prior classical mechanical knowledge [16]. Similarly, Hoehn and Finkelstein investigated the circumstances of when modern physics students used analogies and other ontologies to connect classical and quantum ideas [17]. These papers demonstrate the versatility of analogies towards aiding student learning. Although Wittmann and Morgan focuses on using everyday experiences while Schermerhorn et al. and Hoehn and Finkelstein focus on classical mechanics, all frame those experiences and knowledge in a way to help students learn quantum mechanics.

Beyond the classroom, professional scientists use analogies to help overcome conceptual challenges in their research [18, 19]. The fact that analogies are a tool used in authentic research environments demonstrates the usefulness and importance of integrating them into upper-division curricula to benefit problem-solving skills [18].

However, research has also shown that caution is necessary when using analogies, as appropriate usage (context and phrasing) and sufficient explanation are necessary for the meaning and importance of analogies to be understood by students [13, 14, 20]. Otherwise, students may map unintended features from the base to the target, or even extend the analogy beyond the intend scope, which leads to an inaccurate understanding.

Our study complements prior research by focusing solely on analogy use in textbooks, whereas previous research has predominately focused on classroom implementation. Focusing on textbooks will allow us to potentially learn what bases of knowledge authors assume readers have, what target knowledge readers may learn from analogies, and how analogy use in textbooks may be improved.

III. METHODS

To explore the ways analogies are used in quantum mechanics instruction, we conducted a textbook analysis [21, 22]. The textbooks analyzed include Quantum Mechanics
by McIntyre [23] and Introduction to Quantum Mechanics, Third Edition by Griffiths [24]. These textbooks were selected based on wide usage in physics classrooms and varying writing styles and approaches to teaching.

McIntyre was analyzed first, because the chapters were shorter and contained fewer topics. The chapters in Griffiths covering similar topics were then analyzed. Chapter selection was based on topics typically covered in a first semester quantum mechanics course: operators and measurement, the wave function and time evolution, angular momentum, and identical particles.

We iteratively coded chapters in these textbooks in order to develop themes regarding their analogy usage. During the first pass, the textbook chapters were read and we applied a single a priori code of analogies. Each identified analogy was also coded with a short descriptive initial code (following grounded theory) summarizing the base, target, and context/topic [25]. For this analysis, we use the following definition of analogy: an analogy is a mapping of features from a base of familiar knowledge to features of a target of unfamiliar knowledge [9, 10, 12]. To be coded as an analogy, a base and target domain, along with at least one mapped feature, must be identified; the analogy could be a mapping to show either similarity or dissimilarity of features between the base and target. Identifying analogies was not a trivial process, given that some analogies have implied components. As such, it became helpful to identify key words that indicated a relationship between a base and target. Given the usefulness of these key words, they became a part of the methods used for identifying select analogies. We examine trends in key word use within Sec. IV A.

During a second pass through the data, codes were further categorized by corresponding base. The main categories of base knowledge were Comparison to Classical Mechanics, Mathematical Comparisons, and Everyday examples. An analogy was coded with Comparison to Classical Mechanics when the base of the analogy incited knowledge from classical mechanical classes; Mathematical Comparisons was used when the base of the analogy referenced knowledge from mathematics classes (frequently linear algebra); lastly, an analogy was coded with Everyday examples when the base involved common everyday experiences (for clarification of this term, see Sec. IV B). When applying these main codes, it did not matter whether the analogy was meant to demonstrate similarity or dissimilarity to quantum mechanics.

For example, “The quantum oscillator is strikingly different from its classical counterpart - not only are the energies quantized, but the position distributions have some bizarre features” (Griffiths, Ch. 2 pg. 71) was given the following initial code: CM oscillator to QM oscillator, probabilities. Identification of the base knowledge as a classical oscillator led to further categorization, which we labeled as Comparison to Classical Mechanics. Throughout the coding process and writing of this paper, regular meetings were held between all three authors to establish reliability of the codebook and the emergent categories.

A. Analogies Indicated by Key Words

The discovery of keywords and phrases came about during the analysis, and so was not initially coded for. Instead, key words became tools to identify select analogies and became their own unanticipated research topic.

Determining what is and what is not an analogy is not a trivial process. Thus, it is helpful to pick up on certain key words that indicate a relationship between a base and target. The following is a list of common words found in the analysis that indicate an analogy: is/are, analogue/analogous, just as/like, like/same/similar, satisfy/satisfies, different/difference, etc.

The key words “is/are” are most frequently used to set an equivalence or satisfying mathematical properties. For example, “$A^\dagger$ is the hermitian conjugate of $A$” (Griffiths, Ch. 2 pg. 63) uses a base of linear algebra and matrices to explain the hermitian conjugate of a quantum mechanical operator. The words “analogue/analogous” are mainly used to indicate a mathematical relationship or to identify a similar way of solving certain mathematical systems, such as, “The Schrödinger equation plays a role logically analogous to Newton’s second law” (Griffiths, Ch. 1 pg. 16). “Just as/just like” is used to identify similarity between mathematical properties or solving a mathematical system; “The Schrödinger equation determines $\psi(x, t)$ for all future time, just as in classical mechanics, Newton’s law determines $x(t)$ for all future time” (Griffiths, Ch. 1 pg. 16). Additionally, “like/same/similar” are used most frequently to note similarities between solving techniques of quantum systems and comparing quantum mechanics to linear algebra concepts, such as “You may not know how to solve [the Schrödinger equation], but you do know how to solve a very similar one - Newton’s second law” (McIntyre, Ch.5 pg. 151). Although these keywords and phrases have inherent similar meaning and function, they have different presentations based on the author.

Further keywords include “satisfy/satisfies” and “different/difference”: “satisfy/satisfies” is solely used to identify shared mathematical properties, and “different/difference” is most frequently used to identify specific differences between fundamental systems, especially if there are known similarities between the systems as well.

Still, keywords and phrases are not always helpful in identifying analogies, as the relationship between a base and tar-
get is often implied. Take the following quote for example: “Classical mechanics relies on Newton’s second law \( F = ma \) to predict the future of a particle’s motion. The ability to predict the quantum future started with Erwin Schrödinger and bears his name” (McIntyre, Ch. 3 pg. 68). In this quote, McIntyre uses similar wording (“predict the future” and “predict the quantum future”) to establish a connection between the base of Newton’s second law and the target of the Schrödinger Equation, but there is no cue word or phrase that indicates the analogy. Rather, it is the parallel sentence structure. In a similar manner, Griffiths uses a semi-colon to indicate a contrasting analogy between the previously covered bases of the infinite square well and harmonic oscillator and the target free particle: “Because the infinite square well and harmonic oscillator potentials go to infinity as \( x \to \pm \infty \), they admit bound states only; because the free particle potential is zero everywhere, it only allows scattering states” (Griffiths, Ch. 2 pg. 83).

### B. Connecting everyday experiences to the quantum world

Both McIntyre and Griffiths use analogies in order to make connections between everyday experiences and the quantum world. In this case, everyday experiences refer to occurrences and scenarios that most people (not only scientists) encounter and would recognize. Analogies connecting everyday experiences to quantum mechanics were coded as Everyday experiences. It was found that Griffiths uses this type of analogy more than McIntyre.

Examples in McIntyre of this include the following: “From these plots, it is now clear why we call \( \psi(x) \) the wave function. These energy eigenstates have a ‘wavy’ spatial dependence, much like the modes on a guitar string” (McIntyre, Ch. 5 pg. 124). To emphasize this point, McIntyre continues the analogy: “First, the energy levels can be adjusted, or ‘tuned,’ by changing the thickness of the quantum well layer” (McIntyre, Ch. 5 pg. 147). Through the everyday base of a guitar, McIntyre demonstrates how the mathematics of a quantum potential well work.

In comparison, Griffiths relates the No-Clone Theorem to a Xerox machine: “Indeed, if you could build a cloning device (a ‘quantum Xerox machine’) quantum mechanics would be out the window,” and continues the analogy by saying “schematically, we want the machine to take as input a particle in state \( |\psi\rangle \) (the one to be copied), plus a second particular in state \( |X\rangle \) (the ‘blank sheet of paper’), and spit out two particles in the state \( |\psi\rangle \) (original plus copy)” (Griffiths, Ch. 12 pg. 583). Thus, Griffiths compares the functions of a Xerox machine to that of a hypothetical quantum cloning device as a means to contrast classical and quantum behaviors.

Similarly, McIntyre leverages hypothetically fluctuating sock properties as a means to demonstrate the unintuitive nature of quantum mechanics: “Quantum particles behave as mysteriously as Erwin’s socks - sometimes forgetting what we have already measured” (McIntyre, Ch. 1 pg. 1). This example serves to demonstrate how intuitive ideas about the properties of socks are insufficient for understanding the spin properties of quantum systems.

### C. Leveraging Previous Mathematics Towards Quantum Systems

Another common base of knowledge that emerged from the textbook analysis is mathematical knowledge, as might be seen in other mathematics or non-quantum physics courses. Analogies that used mathematics and mathematical processes as the base knowledge were coded with Mathematical Comparison. This can take one of two forms: using similar problem-solving techniques or using known mathematics to build a basis for concepts, notation, and associated formalism.

Concerning similar problem-solving techniques, McIntyre compares how to solve the differential form of the Schrödinger equation to solving Newton’s second law: “You may not yet know how to solve [the Schrödinger equation], but you do know how to solve a very similar one - Newton’s second law” (McIntyre, Ch. 5 pg. 151). Similarly, Griffiths has the following quote comparing the time evolution between the two: ‘The Schrödinger equation plays a role logically analogous to Newton’s second law: given suitable initial conditions (typically, \( \psi(x, 0) \)), the Schrödinger equation determines \( \psi(x, t) \) for all future time, just as, in classical mechanics, Newton’s law determines \( x(t) \) for all future time” (Griffiths, Ch. I pg. 16).

Often, both textbooks make analogies relating linear algebra concepts (base) to quantum mechanical concepts (target) in order to build the mathematical formalism of quantum mechanics. Thus, both authors assume that readers are knowledgeable of linear algebra, and that readers should be able to understand the analogical connections that the authors make. When forming this relationship between linear algebra and quantum mathematics, both McIntyre and Griffiths use key words, such as “satisfy/satisfies” and “is/are” among others, which may be seen in the following examples.

Griffiths compares abstract vectors (base) with wave functions (target) and linear transformations (base) with operators (target): “Mathematically, wave functions satisfy the defining conditions for abstract vectors, and operators act on them as linear transformations. So the natural language of quantum mechanics is linear algebra” (Griffiths, Ch. 3 pg. 119).

Similarly, McIntyre compares geometric vectors (base) and basis vectors (target): “Continuing the mathematical analogy between spatial vectors and abstract vectors, we require that these same properties (at least conceptually) apply to quantum mechanical basis vectors” (McIntyre, Ch. 1, pg. 11). McIntyre establishes a connection between the linear algebra (base) and quantum (target) versions of the adjoint: “Equation (2.50) tells us that the matrix representing the Hermitian adjoint \( A^\dagger \) is found by transposing and complex conjugating the matrix representing \( A \). This is consistent with the defini-
tion of Hermitian adjoint used in matrix algebra” (McIntyre, Ch. 2 pg. 44).

D. Using classical knowledge to contrast with quantum mechanics

Analogies are frequently used to demonstrate similarities, but analogies can also be used to emphasize differences. For instance, while previous examples demonstrated similarities between classical mechanics and linear algebra in connection to understanding quantum mechanics, analogies can also be used to demonstrate the differences between them. Analogies that contrasted classical mechanics and linear algebra to quantum mechanics were coded under Comparison to Classical Mechanics.

In McIntyre and Griffiths, this difference most often comes in the form of stating a quantum mechanical phenomena or mathematics and discussing how it is different from classical mechanics, corresponding classical intuition, or linear algebra. These classical-quantum analogies frequently discuss experimental results and mathematical expressions.

Experimental results are discussed in both textbooks, and the following are examples from each textbook in relation to the phenomenon of quantum tunneling or barrier penetration. From Griffiths, “Classically, of course, a particle cannot make it over an infinitely high barrier, regardless of its energy [...] Quantum scattering problems are much richer: The particle has some non-zero probability of passing through the potential [...] We call this phenomenon tunneling” (Griffiths, Ch. 2 pg. 114). From McIntyre, “Quantum mechanical particles have a finite probability of being found where classical particles may not exist! This is a purely quantum mechanical effect and is commonly referred to as barrier penetration” (McIntyre, Ch. 5 pg. 133). Both quotes demonstrate how a classical particle cannot pass through an infinite barrier, but a quantum particle can.

The authors also demonstrate differences between mathematical expressions using analogical reasoning. The following are examples from each textbook of the following base-target pairing: comparing the classical use of a complex-valued function (base) to that of quantum mechanics (target). From Griffiths, “Incidentally, in electrodynamics we would write the azimuthal function in terms of sines and cosines, instead of exponentials [...] But there is no such constraint on the wave function” (Griffiths, Ch. 4 pg. 176). From McIntyre, “Note that the imaginary components of these kets are required. They are not merely a mathematical convenience as one sees in classical mechanics” (McIntyre, Ch. 1 pg. 25). Both quotes demonstrate how the wave function in quantum mechanics utilizes imaginary components, whereas in classical mechanics, they are not.

V. CONCLUSION AND DISCUSSION

Our analysis of analogy use in quantum mechanics textbooks reveals two main ideas, in connection to our research questions and results. One, key words or phrases can be helpful in identifying select analogies, and serve to emphasize different types of analogical mappings. And two, analogies have many different roles in quantum mechanics textbooks; analogies can help (1) connect everyday experiences to the quantum world, (2) solve new quantum systems through using prior knowledge from mathematics and classical physics, and (3) understand the differences between classical and quantum mechanics.

Although our analysis focused on textbooks, it suggests areas to investigate around teaching and learning. We observed that each textbook author will make assumptions as to the current level of student knowledge, and will base analogies off that. If the reader’s knowledge is not aligned with the analogy’s base knowledge, the intended target will probably not be understood.

When it comes to using analogies as learning tools, we hypothesize that students may have difficulty recognizing analogies or the extent of the analogical mappings used. We observed that, besides normal variation in writing style (wording, sentence structure, etc.), variation within textbooks may additionally appear within key word usage, specificity in topic coverage, and as a result, frequency and appearance of analogies. For example, Griffiths heavily uses analogies based on everyday experiences, while McIntyre prefers to use analogies for comparing quantum to classical mechanics. Thus, it may be beneficial to give students practice in identifying analogies, and particularly bases, targets, and corresponding mappings. In general, when analogies are designed at an appropriate level, are readily identifiable, and students are given time to understand the analogical mappings, they can be useful tools in aiding student understanding and learning.

Reflecting on the relationship between analogies and models, we viewed mathematical models as a particular type, or subset, of analogies. Specifically, all mathematical models are analogies but not all analogies are mathematical models; the distinction is in the directness and completeness of the mapping [26]. For instance, some examples above may be viewed as mathematical statements or definitions; these are analogies, but depending on the directness of the statement, could also be viewed as mathematical models.

Potential limitations of the current analysis include only focusing on particular topics in the textbooks and not examining key words more closely. These limitations may be addressed in future research. Potential next steps include investigating how students recognize and interpret analogies within texts and in the classroom. More broadly, this research has implications for educating the general public and students.

We would like to acknowledge the following grants for the opportunity to conduct this research: NSF REU 2149957 and NSF DGE 1846321.


Characterizing representational gestures in collaborative sense-making of vectors in introductory physics

Stacy M. Scheuneman and Virginia J. Flood
Department of Learning and Instruction, University at Buffalo, Buffalo, NY, USA

Benedikt W. Harrer
Department of Physics, University at Buffalo, Buffalo, NY, USA

An understanding of vectors and vector operations is crucial for success in physics, as this serves as the foundation for various essential concepts, including motion and forces. Previous research indicates that only a fraction of introductory physics students have a usable knowledge of vectors and vector operations, and that more attention should be given to how students make sense of vectors. We examined classroom video data from an introductory physics course wherein students worked collaboratively through learning activities to introduce vectors and vector operations. During these activities, students' employment of gesture as a representational mode facilitated group sense-making. We propose a preliminary taxonomy of gestures for representing vector magnitudes, directions, and initial and terminal points. By identifying and characterizing the gestures used by students, we can gain insights into their learning processes and conceptual understanding of vectors, which can inform instructional design and teaching practices.
Prior research on introductory physics students’ vector knowledge indicates that the traditional treatment of vectors and their operations is insufficient. Using the results from administration of the Vector Knowledge Test, Knight determined that only 50 percent of students in a calculus-based first semester physics course demonstrated the ability to perform vector addition [1]. Furthermore, many introductory physics students have difficulties in carrying out vector addition and subtraction, even after the completion of a first semester mechanics course. For example, on a graphical vector assessment administered in a second semester, algebra-based introductory physics course, 73 percent of students answered a one-dimensional addition problem correctly, 44 percent answered a two-dimensional addition problem correctly, and 35 percent answered a two-dimensional subtraction problem correctly [2]. Similar findings saw about 50 percent of students able to correctly answer a vector addition problem after a semester of traditional instruction [3].

The findings from an investigation of arrow and algebraic vector notation suggests that while students who are given an algebraic notation assessment of vector addition achieved higher scores, students given an arrow notation assessment answered with solution methods requiring qualitative component canceling and angle reasoning at a higher rate [4]. This may indicate that the arrow notation presents students with a greater challenge due to subtleties in the representation of spatial characteristics such as direction and magnitude. Although arriving at the correct numerical answer is an important goal for physics students and their instructors, developing a conceptual understanding of problems and their solutions is arguably of greater importance. Doing so requires that students are provided opportunities to take advantage of different representational modes, including gestural modes.

Gestures are physical movements that people make using their hands and arms which are meaningful substitutions for ideas and entities. In addition to playing a significant role in communication, there is a growing body of evidence that indicates that gestures affect thinking and learning [5]. The evidence is especially strong for domains in which spatial reasoning is an essential competency, such as the natural sciences and mathematics [6] [7] [8].

Representational gestures are gestures which are visually similar to what they reference. Representational gestures can be further categorized as iconic gestures, which resemble the shape of what is being gestured about, or abstract deictic gestures, in which an empty space in front of the body is pointed to and treated as being occupied by some imaginary entity [9].

Gestures’ ability to convey spatial information has been established as a productive tool in geometric reasoning and graphing of mathematical functions [10] [11]. Due to the highly spatial nature of vectors and vector operations, students’ use of gestures during sense-making activities which are focused on developing these foundational skills should be of particular interest to introductory physics educators.

Both the form of gesture and the nature of vector properties are considered in the characterization of vector gestures. The four main forms of representational gestures are:

- imitation, in which the hands are used to mime actions associated with an object
- portrayal, in which the hands are used to represent an object’s form
- drawing, in which the hands are used to trace or outline a shape
- sculpting, in which the hands are used to mold or carve a shape [12].

The use of dynamic gestures - those which represent change - has also been linked with increased success in mathematics proof development [13]. The nature of vector properties is classified using a schema for organizing spatial skills. This schema considers whether the referenced properties are intrinsic or extrinsic, as well as whether they are dynamic (changing) or static (unchanging) [14]. Intrinsic properties are the spatial features of an object, such as the shape and size. For vectors, the magnitude is an intrinsic property. Extrinsic properties are the locating features of an object, in relation to other objects or to a frame of reference. For vectors, the direction and the initial and terminal points are extrinsic properties.

Despite the difficulties students have with vectors and our knowledge that gestures play an important role in students’ spatial reasoning in STEM, we know surprisingly little about how students use gestures to make sense of and reason together about vectors. The purpose of this qualitative study is to investigate how students gesture about vectors, and how these gestures convey students’ understanding of various vector properties. The goal of this paper is to characterize how students use gesture to make sense of vectors. We discuss the affordances and limitations of these different types of gestures for understanding and reasoning about vectors. Students were observed during collaborative problem solving and sense-making activities while working through learning activities to introduce vectors and vector operations in two dimensions. As part of a larger project to examine physics students’ use of dialogic gesture in sense-making, we propose a preliminary taxonomy of gesture characteristics for representing vector properties, accounting for static and dynamic representations of vector magnitudes, directions, and initial and terminal points.

II. METHODOLOGY

In this qualitative study, a microethnographic approach [15] was used to analyze the gestures of undergraduate physics students. A microethnographic approach allows us to reveal and document unfolding processes of sense making instead of just examining outcomes at fixed points in time. The primary data source is a video corpus of an algebra-based introductory undergraduate physics course using the Collaborative Learning through Active Sense-Making in Physics...
FIG. 1. Two students simultaneously portray the heads of two force vectors acting on the same object in opposite directions.

(CLASP) curriculum [16]. Specific segments analyzed in this study were recordings of three different groups working on three connected activities on vector properties and operations. For each group, 30 minutes of video were analyzed.

In the first activity, students were tasked with determining the range of possible values for two masses when combined, and for two forces when combined, to contrast the differences involved in adding scalar quantities and adding vector quantities. In the second activity, students determined a method for adding vectors tip to tail, first using two vertical force vectors of different magnitudes with the same direction; then extended this method to two vertical force vectors of different magnitudes acting in opposite directions and two force vectors of different magnitude acting perpendicular (one vertical and one horizontal) to one another. In the third activity, students performed a graphical vector subtraction for two diagonal position vectors to find the displacement. Work took place in groups of three to four students. Students were seated together and were directed to show their work on a large whiteboard.

In identifying and characterizing gestures, gestures were grouped together by the form of the gesture (imitation, portrayal, drawing, and sculpting), and the intrinsic/extrinsic and static/dynamic nature is considered for the referent vector properties of direction/orientation, magnitude, and initial and terminal points.

Four classes of representational gestures were identified in this analysis, and specific cases of each are analyzed: portraying, drawing, sculpting, and imitating handling of the vector.

We present our characterization of these different types of gesture in detail to allow the reader to judge the suitability of our categories for describing students’ use of gesture. The first author watched each example repeatedly to establish reliability and accuracy of each characterization. In future work, we plan to test the reliability of these categories using independent coders to measure inter-rater reliability.

III. CHARACTERIZING GESTURES ABOUT VECTORS DURING SENSE-MAKING

During the sense-making activities at the focus of this study, students were observed gesturing about intrinsic and extrinsic properties of the vector which were still being made sense of using dynamic gestures. Conversely, students tended to gesture about characteristics using static gestures when they were more certain (although not necessarily correct) in their understanding.

A. Portraying the vector

One class of gestures for vectors observed is use of the hands to portray the vector. While portraying the vector, the hands are angled at the wrist and the fingers are extended in line with the flattened palm. For example, for the gestures shown in Fig. 1, students portray the heads of two oppositely facing vectors head-to-head. In this gesture, we see static representation of the vectors’ direction and terminal points, as well as the spatial relationship between the two vectors.

Some information about the vectors may be missing from this gesture. It is unclear whether information about the initial points and magnitude is conveyed, since the portrayal of the vectors using the hands is limited by the hands’ symmetry. The student on the left, however, has their lower hand angled straight with their forearm, which may suggest that static information about the initial points and magnitudes is being conveyed. During the activity in which these gestures take place, students are attempting to add two oppositely pointing vectors, with the vector of greater magnitude pointing up.

B. Drawing the vector

Students drew hypothetical vectors by tracing a line by moving an extended finger linearly back and forth between two points, such as shown in Fig. 2. The magnitude and initial and terminal points are represented statically. The static nature of the initial and terminal points in the gesture can be seen as the tracing never passes the boundary of either point. The representation of the direction of the hypothetical vector is dynamic, indicated by the switching of directions of the stroke at least once. Although similar to gestures that portrayed the vector, with these drawing gestures, speakers did not commit to one direction.

These drawing gestures helped the speaker convey information about the orientation and placement of the vector without specifying direction. Often, this occurred when the direction of the resultant vector had not yet been determined by the speaker.

For example, following the gesture made by the student in Fig. 2, another member of the group draws a dashed line without an arrowhead, then asks for clarification of which direction the vector should be pointing multiple times. In re-
FIG. 2. A student traces a hypothetical vector during the graphical vector subtraction activity, in which students need to determine a method for finding the displacement vector from two position vectors.

spose, the gesturing student does not respond by stating a direction but refers back to the reference materials and says they should write out an equation.

C. Sculpting the vector

The two previously described gestures can be combined to convey additional information, in which the speaker sculpts around a vector with a moving, flattened hand, such as that seen in Fig. 3. Like gestures which draw the vector, the direction is represented dynamically while the magnitude and initial and terminal points are represented statically. There is a subtle difference with the addition of the portrayal of the vector as well, with the pointed fingers changing direction and the speaker specifically questioning the direction.

In the activity shown in Fig. 3, students in the group are discussing whether subtraction of the one vector (the subtrahend) from another (the minuend) would require arranging the subtrahend in the same direction as the minuend (shown in the top panels of the figure) or in the opposing direction (shown in the bottom panels of the figure).

D. Imitating handling the vector

When students used gestures to imitate actions done to vectors, they held it as though it were a physical object taking up space in their hands. By rotating or translating the invisible vector while holding the ends of the imaginary vector, which is done with either the fingers of two hands or two fingers on the same hand, or gripping the vector along the middle, the speaker conveys information about extrinsic characteristics of the vector.

When simulating a rotation of the vector, such as in the gesture shown in Fig. 4 (left), the magnitude and initial point are represented statically, with the distance between the hands remaining the same and the left hand remaining in place. Direction and the terminal point are represented dynamically, as seen in the changing position of the right hand. After this gesture, they direct the other member of the group to draw a vector which only shows a reflection over the y-axis, as opposed to the complete 180 degree rotation represented in the gesture. This suggests that although the gesture represents what should be done to the vector, the speaker may be uncertain.

When transforming a vector via translation, such as in the gesture shown in Fig. 4 (right), students simulate holding on to the ends of the vector and sliding it across the diagram. In this type of gesture, the magnitude and direction of the vector are represented statically, while the initial and terminal points are represented dynamically.

In this interaction, the speaker was trying to convince their group member that the direction and magnitude were drawn incorrectly in the diagram by showing that the magnitude did not match the distance between their hands and that the direction didn’t change when they had already agreed that it should be the opposite. After the gesture, the group member erases the additional dashed lines at the end of the vector and changes the direction and placement of the arrowhead.

IV. DISCUSSION

Across the examples of gestures about vectors described in this preliminary analysis, dynamic representations of vector properties accompanied students’ active reasoning with the vector property. For example, when direction was being represented dynamically in a gesture, the speaker was often grappling with figuring out the vector’s direction. Conversely, static representations were accompanied with a lack of active reasoning - for example, speakers who gestured using static representations of direction were not focused on determining direction.

Few students observed appeared to have difficulty in under-
standing the magnitude of vectors, and no dynamic representations of magnitude were observed in this initial exploration of the data. This lines up with the previously mentioned literature on vector understanding of introductory physics students, which shows that many students have an adequate understanding and ability to determine the magnitudes of vectors but struggle with determining their direction.

In the first activity, although many students were able to distinguish between adding scalar quantities and adding vector quantities, student initiated consideration of differing directions beyond parallel and perpendicular combinations of vectors was not observed. This may indicate that many students possess only a rudimentary understanding of vectors having direction when coming into introductory physics.

Similarly, difficulty with managing vectors with different directions was observed in the second activity. Although students were able to add two parallel vectors with the same direction with ease, and decide that a tip-to-tail method was required to do so, students were not able to subsequently apply this method to two parallel vectors with opposite direction and the two perpendicular vectors without struggling with an attempt to add vectors using a tip-to-tip representation.

Gestures representing direction and initial and terminal positions dynamically were commonly seen in the third activity, which coincided with students having greater difficulty in determining the placement and direction of vectors during the graphical vector subtraction.

V. CONCLUSIONS AND FUTURE RESEARCH

By analyzing a small subset of the large corpus of video data in this exploratory study, we were able to identify trends in students’ gestures about vectors, which provided valuable insights into their understanding of vector properties. We observed that students who dynamically gestured about certain vector properties were often in the process of making sense and developing understanding about how those properties were at play in the context of the activity. Conversely, vector properties which were represented statically in gestures tended to be those which did not require as much attention from students.

While these findings provide only an initial look into the relationship between students’ gestures and the evolution of their understanding of vector properties, they open up several avenues for future research. The next step in furthering our understanding of this relationship is to conduct a larger-scale analysis of video data, encompassing a broader range of students, modes of communication and representation, and physics learning contexts, including activities farther in the curriculum which require more challenging applications of vectors.

The limitations of this study include its small sample size and that it provides a close analysis over a short period of time and therefore cannot shed light longitudinally on how students use gestures to think about vectors and how this may impact learning over time. However, this study present first steps for a more comprehensive analysis. Future work could examine how students’ use of gesture changes over time as they learn about vectors and also could look for patterns in larger samples of students.

In conclusion, this exploratory study has established a foundation for understanding the role of gestures in students’ comprehension of vector properties. Our initial observations highlight the potential of gestures as indicators of active sense-making and provide a basis for future research. By expanding the range and depth of investigation, we can deepen our understanding of how gestures can effectively support both the development of vector-related knowledge and skills, as well as sense-making and problem-solving practices in the wider domain of physics.

ACKNOWLEDGMENTS

This work is supported by NSF ECR-EHR Core Research Award No. 2201821 DiGEST Physics: Dialogic gesture in collaborative sense making in physics.
Despite the positive gains towards student learning outcomes and engagement, active learning has been shown to potentially increase student anxiety due to a fear of negative evaluation. A pedagogical strategy proposed to mediate this issue is known as error framing; it asks instructors to encourage a perception of errors as being a natural part of the learning process. Previous work on this project investigated how graduate teaching assistants (GTAs) operationalized error framing during their training in a mixed-reality simulator but did not investigate their usage of it in their classrooms. This analysis characterizes the error framing statements made by GTAs during a set of classroom observations. We find that GTAs who employ error framing effectively avoid statements that might decrease student comfort and instead tend towards implicit, indirect strategies.
I. INTRODUCTION

As active learning has grown in popularity, many people in the STEM education community have realized the importance of training graduate teaching assistants (GTAs) in pedagogical techniques that help them facilitate active learning classrooms [1]. Additionally, researchers and educators have described various strategies and techniques instructors can use to encourage student participation. One such technique is error framing.

Error framing is an instructional technique where an instructor attempts to encourage participation by creating an environment where errors are seen as a natural part of the learning process [2,3]. This technique is theorized to increase participation by reducing student anxiety through reducing the “fear of negative evaluation,” the fear that a student will be judged by their peers and instructors for making a mistake [4,5]. By framing errors as an important part of the learning process, instructors can mitigate this fear and reduce student anxiety. Error framing can occur when an instructor is engaged with a student or as part of the introduction to an activity [6]. For this paper, we will call error framing statements which occur directly after a student’s mistake “direct error framing”, as these statements attempt to directly frame a student’s incorrect answer. We refer to statements made during an introduction or as a general statement as “error climate statements”, as these statements are not preceded by a student error and instead aim to create a climate where students feel comfortable making mistakes.

We argue error framing is an important skill for GTAs as we believe instructors have a role in helping students “save face” when a student publicly offers an incorrect answer. This is rooted in the “facework” framework, where “face” refers to a student’s self-image they hope to present to others [7]. By having a student answer a question publicly, an instructor presents a “face threat,” as an incorrect answer may reflect negatively on the student. However, by error framing, instructors can engage in “facework” meant to protect students from this threat. This positive facework helps make classrooms a safer space for students to engage in risk taking and independent thinking.

In contrast to the established benefits of error framing, in previous work, students identified parts of example error framing statements that decreased their comfort [8]. For example, students disliked when GTAs started by directly commenting on how the student was wrong, focused on the student’s mistakes, hedged when acknowledging the student’s ideas as understandable, or used certain negative language such as “mistake” or “error”. This creates a potential concern that, if used incorrectly, error framing could increase student discomfort instead of decreasing it. Thus, it is important to see how GTAs operationalize error framing while teaching students.

Prior research by the team characterized the error framing statements made by GTAs during their professional development in a mixed reality simulated classroom [8]. However, GTAs’ implementation of trained pedagogical techniques in their classroom varies from how they were trained [6,9]. Thus, it is not sufficient to study how the technique is implemented in training. In this paper we characterize GTAs usage of error framing in their classrooms in comparison to the themes established in prior work.

II. METHODS

A. Participants

This study was conducted at a large, research-intensive university in the southeastern United States. The data set was pulled from a sample of classroom observations of 27 different GTAs over three semesters: spring 2019, fall 2019, and spring 2020. Some GTAs participated in multiple semesters, and some did not, depending on both their teaching assignment and their choice to consent to research in a particular semester. The GTAs taught an introductory physics “mini-studio” with a class size of around 32 students [9]. The mini-studio style classes featured a 75-minute tutorial based on the University of Maryland Open Source Tutorials [10], followed by a 15-minute quiz, and then 80 minutes of an Investigative Science Learning Environment (ISLE) curriculum based lab [11].

GTAs participated in a professional development session where they practiced using various pedagogical techniques, including error framing [8,9,12]. After this training, the GTAs were observed in select classes throughout the rest of the semester. Four classes were observed for each GTA in spring 2019 and three classes were observed in fall 2019 and spring 2020. Observation numbers varied due to hurricanes and the beginning of the COVID-19 pandemic. These observations were conducted using a modified version of the Laboratory Observation Protocol for Undergraduate STEM (LOPUS) [13]. LOPUS is an observational protocol that involves observers recording the occurrence of specific behaviors, such as a student asking the GTA a question, during two-minute intervals. For our observations, we added an additional code for verbal feedback that contained error framing [14]. Of the 27 GTAs who participated in the professional development during some of the three semesters, 14 used error framing during their observations and a total of 39 two-minute intervals were coded as containing error framing.

B. Positionality

Researchers working on this project were involved in various other aspects of the mini-studios. JC was involved in the curriculum development of the mini-studio labs and
sometimes led GTA prep. CD and TW both at various times led GTA prep meetings during these semesters and conducted some of the observations considered. DS participated in this training in a later semester and taught a different version of this course as a GTA.

C. Themes and Codes

Prior research by members of the team established common themes related to the implementation of error framing [8]. These themes and codes were utilized for this analysis. One set of themes, displayed in Table I, emerged from the research team’s interpretation of error framing. A second set of codes, displayed in Table II, emerged from an analysis of student reactions to exemplar error framing statements during interviews.

The researcher-generated codes describe error framing statements in two ways: error indication and framing. Error indication describes how a GTA comments on a student’s incorrect response and has two potential codes: explicit and implicit. Framing contains three codes: natural, beneficial, and positive acknowledgment. These codes describe how a GTA attempts to frame incorrect student responses in a positive light. A statement would be coded as natural if a GTA framed a mistake as being natural and common. It would be coded as beneficial if the GTA framed that making the mistake was beneficial for learning. If the GTA acknowledged an incorrect response with praise but did not explain why it was considered positive acknowledgment.

The student inspired codes contain two themes: framing and tone. The framing codes describe student perceptions of ways GTAs frame errors and how it impacts the student’s comfort. This includes framing methods that students thought negatively impacted their comfort, which was not considered in the researcher defined codes. It should be noted the student generated codes for framing are different from the researcher defined codes. This is expected, as the undergraduate students interviewed were not formally introduced to the concept of error framing.

The other theme identified from the student interviews was tone. Tone describes how students perceive the diction and inflection used by the GTA and how it impacts their comfort. For this paper, tone was not considered for a few reasons. First, tone is more subjective than the other codes considered; students may hear the same statement and perceive the tone differently. Additionally, perception of tone depends on context that is unavailable in our data. We only recorded audio files of the observations, so actions such as gestures, facial expressions, and posture, which influence perception of tone, could not be considered for our coding. Due to our inability to satisfactorily code tone, it was not considered in this analysis.

D. Identifying and coding error framing statements

Error framing statements were identified using the time intervals marked as containing error framing feedback from the LOPUS observations. If the observation had an associated audio file, the flagged time interval was then transcribed by hand for further analysis. This resulted in 31 segments. Two coders initially analyzed these segments using the codes from Table I, and discussed their coding with a third researcher, who had conducted some of the observations. The research team noticed that some of the statements identified by observers did not align with the error identification codes. For example, after no students volunteered to answer a question, a GTA said, “It’s okay if you’re wrong. There’s, like I told you, there’s no wrong answer, just focus on what you think.” While this fits our definition of an error climate statement, since it does not address a student’s incorrect answer, it did not match the error indication codes. Thus, we decided to exclude error climate statements from this coding analysis. Two coders coded the remaining statements with the additional framing

TABLE I. Researcher defined codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error Indication</td>
<td>Directly comment that the answer is incorrect using words such as “incorrect” and “misconception”</td>
</tr>
<tr>
<td>Implicit</td>
<td>Avoid a direct comment that the answer is incorrect</td>
</tr>
<tr>
<td>Natural</td>
<td>Frame errors as natural or common</td>
</tr>
<tr>
<td>Beneficial</td>
<td>Frame errors as useful for learning</td>
</tr>
<tr>
<td>Positive Acknowledgment</td>
<td>Use positive words to acknowledge student contribution without elaborating on why the contribution is valuable</td>
</tr>
</tbody>
</table>

TABLE II. Student inspired framing codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledge idea as natural and sensible</td>
<td>GTA acknowledges student’s reasoning as natural or sensible even though it is not correct</td>
</tr>
<tr>
<td>Acknowledge sensemaking effort</td>
<td>GTA acknowledges student’s effort to make sense of the disciplinary content</td>
</tr>
<tr>
<td>Acknowledge idea as common</td>
<td>GTA points out the idea is commonly held by people</td>
</tr>
<tr>
<td>Acknowledge a learning opportunity</td>
<td>GTA makes it known that it is a learning opportunity for everyone</td>
</tr>
<tr>
<td>Provide explanation to subject matter</td>
<td>GTA explains the disciplinary content</td>
</tr>
</tbody>
</table>

Positive Impact on Student Comfort

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Impact on Student Comfort</td>
<td>GTA starts with a direct comment that student’s idea is not correct</td>
</tr>
<tr>
<td>Focus on the error</td>
<td>GTA (although unintentionally) focuses on the fact that students made an error</td>
</tr>
<tr>
<td>Use hedging</td>
<td>GTA uses hedging, such as “kind of”, when acknowledging student reasoning as sensible</td>
</tr>
<tr>
<td>Use negative words</td>
<td>GTA uses negative words, such as “misconception”, when referring to a student’s idea</td>
</tr>
</tbody>
</table>

Negative Impact on Student Comfort

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start with a direct comment of idea being incorrect</td>
<td>GTA starts with a direct comment that student’s idea is not correct</td>
</tr>
<tr>
<td>Acknowledge a learning opportunity</td>
<td>GTA makes it known that it is a learning opportunity for everyone</td>
</tr>
<tr>
<td>Provide explanation to subject matter</td>
<td>GTA explains the disciplinary content</td>
</tr>
</tbody>
</table>
codes from Table II. They discussed their coding with the third researcher and came to agreement. Then, a fourth researcher, who was also part of the observation team, conducted a peer review of the coding for all 31 statements. The research team discussed coding until consensus was reached.

III. FINDINGS

Of the 39 times error framing was coded during the observations, 31 were available for additional analysis because the GTA had consented to their audio file being used for research. After review of these 31 statements, we interpreted 13 as direct error framing, 7 as error climate statements, and 11 as not error framing. These 11 statements, while initially coded by the observers as error framing, ultimately did not match with our definition upon further review. Because the LOPUS observations were made live with no opportunity to review, it is not surprising that some flagged statements did not fit the error framing codes upon further review. Often these statements contained similar ideas to error framing but lacked key aspects. For example, one flagged statement that was ultimately coded as not error framing was “Correct, and that’s a big lesson to learn right now. When we’re talking about weight, you need to be clear are we talking about mass or we talking about actual weight.” Here, the GTA highlights a potentially confusing topic after a student asks him a question. However, the GTA is not responding to a student’s mistake. Instead, the GTA is acknowledging the validity and importance of the student’s answer. While providing this sort of validation to students is similar to the goal of error framing, ultimately this quote features a different behavior than what we want to investigate.

Coding with the researcher-defined and student-inspired codes for the 13 direct error framing statements is shown in Table III. Trends among the data are explained below.

A. GTAs avoid explicit error indication

Our sample of direct error framing statements were overwhelmingly coded as implicit for error indication. Only 1/13 of the direct error framing statements was identified as an explicit error indication. It is possible GTAs were not comfortable with telling a student their answer was incorrect. Prior research found that students dislike when GTAs directly refer to and focus on their mistakes [8]. It appears that GTAs, having been students themselves, tend to avoid statements that create uncomfortable situations for the students. Prior literature describes the debate about how and when to tell a student they are incorrect [15]. Future work will focus on investigating why most GTAs did not explicitly mention a student’s idea was incorrect and how the direct indication of an error impacts students’ feelings and their learning.

B. GTAs avoid student identified negative codes

Prior research identified four potential framing strategies that students identified as having a negative impact on their comfort level. Interestingly, all four of these negative-impact codes had a low occurrence in our sample. Only three of the 13 direct error framing statements aligned with codes which decreased student comfort, as shown in Table III. In two of those statements, GTAs used hedging language. In the other statement, a GTA used negative words. The overall low occurrence rate of these negative codes again indicates that GTAs likely understand and empathetically know what might make a student uncomfortable and try to avoid doing so.

C. GTAs refer to their own mistakes while error framing

A common strategy we saw repeated by many of the GTAs was a strategy we refer to as “GTA self-identification”. This is where a GTA attempts to frame a student’s error by saying that they have made similar or the same mistake in the past. For example, a GTA teaching a circuits lab said, “So you have accidentally flipped both of your leads. That’s okay. Every time I build a circuit, I also flip both of my leads, even though I’m supposed to be good at this.” Not only does this normalize the student’s mistake, but it also presents an opportunity for the GTA to build rapport. An instructor’s ability to be relatable or perceived as a “real person” has been correlated with positive aspects of facework (working to protect or maintain a student’s social identity) [16] which suggests this style of error framing might increase student comfort. Also, when GTAs discuss how they have struggled with the material or concept, students might be encouraged to feel comfortable with their own mistakes.

D. Error climate statements

Despite not being considered in this analysis, our results show that GTAs tend to make error climate statements frequently when using error framing. Out of the 31 statements transcribed, 7 were error climate statements. These statements were often made in response to student confusion about a topic. As an example, in response to a student’s weary sigh, one GTA responded, “It’s okay, you’re supposed to be confused, it’s going to take a little bit to get through this.” In this example, we see that when the student experienced a moment of frustration or confusion, the GTA tried to normalize that experience. Error climate statements could help GTAs address situations where students are so confused they cannot even provide an incorrect answer. In these moments students are likely frustrated, and error framing may help decrease negative student experiences.
TABLE III. Findings for frequency of codes, with exemplar quotes

<table>
<thead>
<tr>
<th>Code</th>
<th>Times Observed</th>
<th>Exemplar Quote</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Researcher Defined: Error Indication</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit</td>
<td>1</td>
<td>&quot;The conceptual process was right, but the value was wrong, but that's okay.&quot;</td>
</tr>
<tr>
<td>Implicit</td>
<td>12</td>
<td>&quot;Right, so we would think that, yeah, but, and in both cases we're gonna do the exact same amount of work.&quot;</td>
</tr>
<tr>
<td>Natural</td>
<td>10</td>
<td>&quot;So, you just had something a little askew. That happens you're not used to building circuits.&quot;</td>
</tr>
<tr>
<td>Beneficial</td>
<td>2</td>
<td>&quot;No, no, but this is like not a thing about physics like this is about something with our instrumentation or measurement. See and this is actually a good lesson about real world stuff.&quot;</td>
</tr>
<tr>
<td>Positive Acknowledgement</td>
<td>3</td>
<td>&quot;The conceptual process was right, but the value was wrong, but that's okay.&quot;</td>
</tr>
<tr>
<td>Acknowledge idea as natural and sensible</td>
<td>8</td>
<td>&quot;Oh. Okay I see where you're coming from, I see because of the word gravity.&quot;</td>
</tr>
<tr>
<td>Acknowledge sensemaking effort</td>
<td>7</td>
<td>&quot;You just switched your voltage and leads, that's very common. All right try again.&quot;</td>
</tr>
<tr>
<td>Acknowledge idea as common</td>
<td>3</td>
<td>&quot;I would like at least two variables on the graph, it looks like this is just your resistance but that's ok, you won't lose points for it because we're learning.&quot;</td>
</tr>
<tr>
<td>Acknowledge a learning opportunity</td>
<td>3</td>
<td>&quot;Usually that's what people think of, when I'm talking to you and I say take into account gravity, that's exactly what you're going to do, you write down FG. This is a problem with the English language, we're actually talking about gravity as a constant, as something that's pulling us, as an acceleration.&quot;</td>
</tr>
<tr>
<td>Provide explanation to subject matter</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Start with a direct comment of idea</td>
<td>0</td>
<td>***</td>
</tr>
<tr>
<td>being incorrect</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Focus on the error</td>
<td>2</td>
<td>&quot;I guess you could say that I just wouldn't...&quot;</td>
</tr>
<tr>
<td>Use hedging</td>
<td>1</td>
<td>&quot;This is a very common mistake.&quot;</td>
</tr>
<tr>
<td>Use negative words</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Student Inspired: Framing**

<table>
<thead>
<tr>
<th>Code</th>
<th>Times Observed</th>
<th>Exemplar Quote</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Researcher Defined: Framing</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| **Frequency of codes, with exemplar quotes**

IV DISCUSSION

In this analysis, we characterized how GTAs implemented error framing in their classes. Through this characterization we noticed GTAs who used error framing statements tended to avoid statements that might decrease student comfort. GTAs may be aware of the impact their statements have on student comfort and avoid statements they think could have a negative impact. Thus, any potential negative impact of error framing statements seems to be moderated by the GTAs themselves. This indicates that error framing may be an effective method for reducing student anxiety, and GTAs should be trained to use it in their classes.

We also observed GTAs use a wide variety of different kinds of error framing statements. We believe it is important that GTAs are trained in how to use these different forms of error framing. GTAs tend to use pedagogical skills only if they align with their own perceptions of the learning process [9]. By teaching GTAs many different forms of error framing, GTAs can likely find a form that they will use.

Future work should explore the impact of error framing when enacted by faculty instructors rather than GTAs. Students have reported GTAs to be more relatable [17] and better at guiding group discussion [18] than faculty. It is possible that faculty use of error framing statements may help to break down the power dynamic between student and instructor, allowing for open communication [19]. On the other hand, faculty may have a more difficult time implementing error framing without negatively impacting student comfort, since students are typically more comfortable with GTAs from the start.

V. LIMITATIONS

There are a few factors which limit the generalizability of our claims. Notably this study only investigated a single style of class and only featured GTAs who chose to use error framing in their classes. While GTAs who chose to implement error framing implemented it well, no claims can be made about the GTAs who did not use the skill. It is likely that if they were coerced to use error framing their implementation would differ from what was observed in this study. Future work will explore why some GTAs did not use error framing. Additionally, live-coding observations is difficult, and it is possible that some error framing statements were not flagged during the observation, which might introduce bias as to what error framing statements were available for analysis.

ACKNOWLEDGMENTS

This work is supported in part by National Science Foundation Award #1725554.


Analyzing students’ assumptions to varying degree of prompting during problem solving

Amogh Sirnoorkar and James T. Laverty
Department of Physics, Kansas State University, Manhattan, Kansas -66502

Reports on pedagogical transformations have called for promoting authentic knowledge-building practices in science classrooms. Making assumptions is one such practice that is integral to “doing” physics. In this study, we analyze the nature and characteristics of students’ assumptions when they are (i) not prompted, (ii) prompted explicitly at the beginning, and (iii) prompted at the end of physics problems. Preliminary observations indicate that students seldom generate assumptions unless prompted. When explicitly asked at the beginning of problem solving, students perceive making assumptions as a separate task dissociated from the problem-solving process. However, when asked to reflect on the validity of their solutions in light of their assumptions, not only do students make assumptions that are closely “tied” to their solutions, but go an extra mile by articulating the implications of the violations of their assumptions. Implications of these findings for instruction and assessment design are discussed.
I. INTRODUCTION

Reports on pedagogical reforms in higher education have called for shifting the focus of classroom learning from rote-memorization of concepts to engaging students in authentic knowledge-building practices [1]. Consequently, researchers have promoted the use of multiple representations [2], argumentation [3], evidence-based [4] and analogical reasoning [5] in physics learning environments. Making effective assumptions is another such component which is integral to “doing” physics. Assumptions dictate the underlying conditions in which a model or principle holds true, and play a key role in idealization and approximations in physics.

Given this significance, researchers have explored how students and professional physicists reason about assumptions. These studies have noted students’ difficulties in converting an “ill-defined” problem into a “well-defined” one through appropriate assumptions [6] and conflating assumptions with algebraic denotation of physical quantities [7, 8]. Recent studies have also noted the centrality of assumptions in theorists practices especially in checking the mathematical behavior of models and their role in troubleshooting [9–12]. Despite its centrality in physics, studies towards reliably eliciting assumptions from students tend to be relatively scarce.

The current manuscript seeks to bridge this gap by investigating how introductory students generate assumptions to varying degrees of prompting in physics problems. We analyze the nature and characteristics of students’ assumptions when they are (i) not prompted, (ii) prompted explicitly at the beginning, and (iii) at the end of physics problems. Preliminary observations indicate that students seldom generate assumptions unless prompted. When explicitly asked at the beginning of problem solving, students perceive making assumptions as a separate task dissociated from the problem solving process. However, when asked to reflect on the validity of their solutions in light of their assumptions, not only do students make assumptions that are closely “tied” to their solutions, but go an extra mile by articulating the implications of the violations of their assumptions. These results suggest the need to explicitly emphasize (i) assumptions in physics pedagogy, and (ii) the “reflection of results” in light of the underlying conditions during problem solving.

This manuscript is organized as follows: In the next section, we discuss theory of assumptions before discussing the physics problems and methods of data collection and analysis in Section III. We then present the results in Section IV before discussing their implications in Section V.

II. ASSUMPTIONS

Researchers across multiples disciplines including business studies and linguistics have contemplated the characteristics and significance of making assumptions. In science education literature, discussions on assumptions have revolved around two broad themes: (i) what counts as an “assumption” and (ii) the purposes served by assumptions. According to Scates [13]: “An assumption is a mental datum which is not fully established, but which is used as a basis for continuing the thought or study.”

Scates extends this definition in arguing that assumptions cover the terrain of facts, principles, or other concepts, “the truth of which is taken for granted for particular purposes without insistence upon specific proof”.

Assumptions have also been described in terms of the two purposes they serve: in making the problem tractable [6, 14] (which we refer as “constraining assumptions”), and in relation to the things that the person making assumptions believes to be true in a problem (“given assumptions”) [9]. One of the ways constraining assumptions work is by making an “ill-defined” problem “well-defined”. Ill-defined problems are the tasks with no specific starting/ending points and with no established process of solving them [15]. Tasks such as modeling an abstract system and designing experiments to validate a theoretical idea represent examples of ill-defined problems. Well-defined problems (e.g., end-of-the-chapter physics problems) on the other hand, are the problems with a specific beginning/ending points and an established laid out process between the two. Typically transitioning from an ill-defined problem space to a well-defined one is facilitated through approximations and idealizations in physics. On the other hand, given assumptions represent the features of a system or model that practitioners consider to be true. Examples include: the absence of friction and the constant magnitude of the acceleration due to gravity during free-fall of an object. We refer to these forms of assumptions which are often taken to be true in a problem as “given assumptions”.

In the current study, we adopt the above-mentioned Scates’ definition of an assumption. Since the problems discussed in this study are well-defined, the purpose that assumptions serve is to mainly reflect what students consider to be true in their approach. In the rest of this manuscript, we address the research question: How does prompting in tasks affect students’ articulation of assumptions during problem solving?

III. METHODS

The objective of the current study is to investigate the characteristics of the students’ assumptions in response to the varying degree of prompting in tasks. The tasks involved real-world contexts based on two different amusement ride contexts: Gravitron and Roller coaster. These contexts were chosen since real-world scenarios provide opportunities to make subjective assumptions during problem solving [6]. The “Gravitron task” consists of a cylindrical rotating ride in which riders lean against a wall. The task asks students to determine whether a rider would fall off the Gravitron’s walls under the specified parameters. Figures 1 and 2 represent the problem statements of the task. Due to the space constraints, we refrain from providing detailed solutions and refer readers to our earlier work [16] for the same.
You are asked to design a Gravitron for the county fair, an amusement park ride where the enters a hollow cylinder, radius of 4.6 m, the rider leans against the wall and the room spins until it reaches angular velocity, at which point the floor lowers. The coefficient of static friction is 0.2. You need this ride to sustain mass between 25-160 kg to be able to ride safely and not slide off the wall. If the minimum \( \omega \) is 3 rad/s, will anyone slide down and off the wall at these masses? Explain your reasoning using diagrams, equations and words.

**FIG. 1.** Statement of the open-ended Gravitron problem

The second problem, the “Roller coaster task”, entails testing of an amusement train ride along a specified path. Students are asked to determine whether the applied force would bring the train to a halt at the end of its track. Figure 3 represents its problem statement along with the given diagram of the track. One of the ways to approach this problem is by determining the difference in potential energies of the train between the points A and D (assuming that the train’s mechanical energy is conserved). This difference in the potential energies will be equal to the train’s kinetic energy at point D. Extracting the value of the train’s velocity from the kinetic energy term, and using the kinematic equation \( v^2 - u^2 = 2as \) (where \( s = 113 \) m), one can calculate the value of acceleration and thus force required to stop this train. Comparing the calculated force with the given value, one observes that the applied force to be inadequate in stopping the train.

The Gravitron task had two versions: an open-ended and a scaffolded version. The open-ended version (Figure 1) contained no explicit prompt asking students to make assumptions. The scaffolded version however, contained three sub-parts with the first sub-part explicitly asking “What assumptions do you need to make to be able to solve this?”. The Roller coaster problem, on the other hand, was an open-ended task but with an explicit prompt at the end asking “Under what conditions do you think your conclusion is valid?” In this way, we had three tasks with varying degree of prompting towards making assumptions: first task having no explicit prompt, second task with an initial explicit prompt to articulate assumptions before solving the problem, and the third asking students to specify the assumptions in which their conclusions would be valid.

On a side note, all three problems were part of a study aimed at analyzing students’ engagement on the scientific practices [1] and were designed using the Three-Dimensional Learning Assessment Protocol (3D-LAP) [17]. While the open-ended Gravitron task was designed to engage students in the practice of “Developing and Using Models”, the scaffolded Gravitron and Roller coaster tasks were designed to elicit the practice of “Using Mathematics”.

Data were collected for the three tasks from introductory students through a series of think-aloud interviews during Spring-2018 and Spring-2019. Data for the first task included responses from 10 introductory students collected in Spring 2018. Of the ten interviews, 2 had audio/video issues and are not part of this study. Data for the scaffolded version of the Gravitron and Roller coaster tasks involved another ten introductory students and the responses were collected in Spring of 2019. We highlight that the data set across the second and the third problem is the same, i.e., the same students attempted the two problems (Figures 2 and 3) in a single think-aloud interview. The interview protocol involved asking students to consider the problem-solving exercise as an untimed exam and to articulate their thoughts out loud. During moments of prolonged silences, the interviewer interjected with questions such as “what are you thinking” to nudge students to continue articulating their thoughts. Students were compensated with $20 for their participation.

We analyzed the interviews by taking into account students’ verbal arguments and their written solutions. For the current study, we focus on the explicit assumptions articulated by students either in their verbal arguments and/or in their written solutions. We particularly focused on the frequency, and the circumstances in which explicit assumptions were invoked or reiterated. The “explicit” assumptions in students’ verbal and written solutions were identified by noting phrases such as “assuming”, “given that”, etc. While the number of assumptions invoked by a participant (frequency) reflected the influence of prompting, the accompanying circumstances reflected on how students employed assumptions during problem solving.

**IV. RESULTS**

In what follows, we describe how students generated assumptions to the varying degree of prompting across three problems represented in Figures 1, 2 and 3.
Engineers are testing a new roller coaster-like ride before it starts functioning. The sandbags are strapped into the train to simulate passengers and the total mass of the train with sandbags is 1000 kg. It is supposed to start from rest at point A and stop at point E. The train starts braking at point D so that it will come to a stop at point E. If the brake system applies an average force of 6749 N, will it be enough to stop the train at point E? **Under what conditions do you think your conclusion is valid?** The heights from the ground to points A, B, C, D and E are 173, 145, 124, 95 and 95 (in m), respectively. The distance from D to E is 113 m.

**FIG. 3.** Statement of the open-ended Roller coaster problem. Part of the problem statement asking to generate assumptions has been highlighted in bold. However the bold font was not presented to students.

**A. Students’ assumptions on open-ended Gravitron task**

The open-ended version of the Gravitron task had no explicit prompt asking students to make assumptions. Out of eight students, we found only one participant articulating their assumptions while solving the problem. At the outset, this finding implies that when not explicitly prompted, articulating assumptions is often not the primary focus of students during problem solving. Our finding is in agreement with existing observation that explicit prompting is often not the primary focus of students during problem solving. This finding implies that when not explicitly prompted, articulating assumptions is often not the primary focus of students during problem solving. At the outset, this finding implies that when not explicitly prompted, articulating assumptions is often not the primary focus of students during problem solving. Furthermore, all the eight participants who articulated assumptions did so only after completing the problem (in response to the prompt) but not during problem solving. This observation reinforces our observation made in the open-ended Gravitron task (Section IV A) that generating assumptions is often not a primary focus of students during problem solving. Also, a key characteristic feature of the student generated assumptions in this task is that few of the participants articulated additional arguments discussing the implications of their assumptions’ violations. For example, one of the participants (Participant 1 in Table I) wrote in their solution:

> “Assuming there is a friction offered by the track itself and is not an ideal system. The remaining [energy] will be dissipated”.

The participant further goes on to mention verbally

> “If you don’t take out that [friction], its gonna create problem.”

This reflection on the violation of the assumptions despite the problem statement not asking for it is an interesting feature in our observations. It is interesting because the same participants did not ponder on the violation of their assumptions while engaging with the scaffolded version of the Gravitron task (Section IV B). These arguments indicate that asking students to introspect the validity of their results in light of the accompanying conditions to be a relatively effective approach in eliciting assumptions during problem solving.

**B. Students’ assumptions on scaffolded Gravitron task**

Unlike the open-ended version, the scaffolded version of the Gravitron task consisted of an explicit prompt asking students to first articulate their assumptions before constructing free body diagrams and solving the problem (Figure 2). All ten participants articulated assumptions which mainly spanned across the orthogonality between the ground and the walls of the Gravitron, riders’ weight, and the coefficient of friction offered by the wall. The generated assumptions have been summarized in the second column of Table I.

Furthermore, approximately half of the total participants generated more than one assumption explicitly and interestingly, none referred to the assumptions while solving the remaining parts of the problem. Based on these observations, we infer that explicit prompting in the initial phases of problem solving can nudge students to generate assumptions but also can make the participants treat the exercise as a separate task dissociated from the problem. This finding substantiates existing observations that explicit prompting in tasks often gets treated as a separate task and leads to the “algorithmic approach” during problem solving [18].

**C. Students’ assumptions on open-ended Roller coaster task**

The third task – Roller coaster problem – was open-ended and asked students to reflect on the conditions in which their solutions were valid (Figure 3). Of the ten participants, eight articulated explicit assumptions in response to the prompting. Of the remaining two, one participant did not complete the problem and the other did not articulate assumptions despite completing the task (summarized in the third column of the Table I). These observations reflect a couple of limitations of this form of prompting. Firstly, asking students to reflect on the conditions that determine the validity of their solutions would force students to employ assumptions after instead of during problem solving. Secondly, successful completion of the task becomes the prerequisite to generate assumptions.

Furthermore, all the eight participants who articulated assumptions did so only after completing the problem (in response to the prompt) but not during problem solving. This observation reinforces our observation made in the open-ended Gravitron task (Section IV A) that generating assumptions is often not a primary focus of students during problem solving. Also, a key characteristic feature of the student generated assumptions in this task is that few of the participants articulated additional arguments discussing the implications of their assumptions’ violations. For example, one of the participants (Participant 1 in Table I) wrote in their solution:
strategy for guiding students in making assumptions.

which their solutions are valid proves to be a relatively effective strategy in eliciting assumptions.

(iii) encouraging students to reflect on the conditions under which their solutions are valid proves to be a relatively effective strategy for guiding students in making assumptions.

students to reflect on the conditions in which their solutions were valid (Figure 3). We observe that students rarely generated assumptions in the first task. However, in the second task, all participants generated assumptions, although they did not effectively integrate them into their problem-solving process. In the third task, not only were students’ assumptions closely linked to their solutions, but they also considered the implications of violating those assumptions.

Based on these observations, we claim that (i) students typically do not prioritize making assumptions explicit when engaged in problem-solving, (ii) explicit prompting to generate assumptions at the beginning of problem-solving tasks may not yield productive results, as students tend to perceive it as a separate task detached from the problem-solving, and (iii) encouraging students to reflect on the conditions under which their solutions are valid proves to be a relatively effective strategy for guiding students in making assumptions.

For instructors, these results indicate the need to emphasize the role of assumptions in classroom instruction, especially by reflecting on the validity of results in light of the underlying conditions. Our findings also provide insights on developing assignments and examinations that effectively nudge students in productively engaging with assumptions. For researchers, our findings call for expanding research on students’ generation of assumptions to various degrees of prompting by focusing on “ill-defined” problems. Explorations probing characterizations of students’ explicit vs implicit assumptions also is a potential avenue.

However, claims made in this study accompany few limitations. Ideally we expect students to blend assumptions while making sense of physics problems. None of the prompting strategies achieved this objective. In response to this shortcoming, we rephrased our third claim that nudging students to reflect their solutions in light of the accompanying conditions was a relatively effective strategy in eliciting assumptions. Secondly, our observations are drawn from responses of only ten introductory students. Observations from larger data-set by taking into account students’ demographics would undoubtedly enrich the results. Three, by the very design, the problems were relatively well-defined. Consequently, our analysis captured the “given assumptions” without exploring students’ “constraining” ones (Refer Section II). Analyzing data from ill-defined problems with varying degree of prompting would capture the spectrum of reasoning around students assumptions in physics.

Future work would involve expanding our analysis to more well-defined problems based on real-life from our data set. We further seek to explore the patterns (and potentially hierarchies) in the students’ generated assumptions in response to varying degree of prompting.

V. DISCUSSION, CONCLUSION, AND FUTURE WORK

We analyzed how students generated assumptions to varying degree of prompting in physics problems. The first problem was open-ended with no explicit prompting (Figure 1). The second problem was the scaffolded version of the first task with an initial explicit prompt to generate assumptions (Figure 2). And lastly the third task was open-ended asking students to reflect on the conditions in which their solutions were valid (Figure 3). We observe that students rarely generated assumptions in the first task. However, in the second task, all participants generated assumptions, although they did not effectively integrate them into their problem-solving process. In the third task, not only were students’ assumptions closely linked to their solutions, but they also considered the implications of violating those assumptions.

Based on these observations, we claim that (i) students typically do not prioritize making assumptions explicit when engaged in problem-solving, (ii) explicit prompting to generate assumptions at the beginning of problem-solving tasks may not yield productive results, as students tend to perceive it as a separate task detached from the problem-solving, and (iii) encouraging students to reflect on the conditions under which their solutions are valid proves to be a relatively effective strategy for guiding students in making assumptions.

For instructors, these results indicate the need to emphasize the role of assumptions in classroom instruction, especially by reflecting on the validity of results in light of the underlying conditions. Our findings also provide insights on developing assignments and examinations that effectively nudge students in productively engaging with assumptions. For researchers, our findings call for expanding research on students’ generation of assumptions to various degrees of prompting by focusing on “ill-defined” problems. Explorations probing characterizations of students’ explicit vs implicit assumptions also is a potential avenue.

However, claims made in this study accompany few limitations. Ideally we expect students to blend assumptions while making sense of physics problems. None of the prompting strategies achieved this objective. In response to this shortcoming, we rephrased our third claim that nudging students to reflect their solutions in light of the accompanying conditions was a relatively effective strategy in eliciting assumptions. Secondly, our observations are drawn from responses of only ten introductory students. Observations from larger data-set by taking into account students’ demographics would undoubtedly enrich the results. Three, by the very design, the problems were relatively well-defined. Consequently, our analysis captured the “given assumptions” without exploring students’ “constraining” ones (Refer Section II). Analyzing data from ill-defined problems with varying degree of prompting would capture the spectrum of reasoning around students assumptions in physics.

Future work would involve expanding our analysis to more well-defined problems based on real-life from our data set. We further seek to explore the patterns (and potentially hierarchies) in the students’ generated assumptions in response to varying degree of prompting.

VI. ACKNOWLEDGEMENTS

Thanks to Dean Zollman for his constructive feedback. This material is based upon work supported by the National Science Foundation under Grant No. 2013339.


Investigating how students engage with a digital planetarium

Alexander Sivitilli
Department of Astronomy, University of Cape Town, Rondebosch, South Africa, 7701

Saalih Allie
Department of Physics, University of Cape Town, Rondebosch, South Africa, 7701

Modern planetariums are digital immersive facilities that promise new teaching opportunities beyond their historic limitations as analog instructional media in astronomy education. We investigated the use of a planetarium as a teaching and learning space by analyzing student responses gathered from a cohort of university students during two separate planetarium visits which constituted part of their introductory astronomy course. An instrument was first developed to probe how the students responded to the overall planetarium experience and how they engaged with the educational content. In the second student visit, show content was specifically designed by us and the instrument was modified based on the results of the first visit. Individual student responses were coded for higher-level categories and concepts. From these concepts we developed a localized mid-level model of student engagement. This led to the notion of a "spectrum of attentiveness" that strongly influences how students engage with relevant educational content.
I. INTRODUCTION

Digital planetariums are immersive facilities with a multitude of capabilities beyond their previous analog limitations. As instructional tools formerly dedicated to simulations of the night sky for astronomy education, new and upgraded domes allow the display of virtually any media able to be rendered by a computer. With these new capabilities, there are many things that can be presented along with the many novel ways that an audience can engage with the space. This is an important consideration as the planetarium is still often used as a teaching tool, especially in astronomy education[1].

Despite its century-long history and the new trend of upgrading planetariums to digital form, existing studies do not reveal established models on how to shape the facility as an effective teaching and learning space. We therefore set out to characterize the teaching and learning space of a digital planetarium by raising two manageable Guiding Questions: GQ1) How do students engage with a digital planetarium? and GQ2) What shapes the teaching and learning space in the planetarium? The present work reports on a sub-component of the investigation that focused on GQ1. This was done through exploring how students engage with a digital planetarium during its use as part of a formal introductory astronomy curriculum at a university. GQ2 was later addressed through the analysis of observational data compiled throughout the investigation, which is outside the scope of this paper.

The investigation’s broad approach to an apparently complex and novel space required a suitable means of systematic exploration. This was done using the Grounded Theory Method (GTM). The GTM exists in a few different flavors with different styles of implementations and underlying philosophies. For our work, we chose the GTM of the pragmatic and constructivist tradition influenced by the work of Charmaz[2] and Bryant[3]. The methodology behind this flavor provided us the means to take account of the complex space and the various social interactions within it, including the researchers’ own role and preconceptions. The end product of GTM is a model grounded in the data, i.e., a substantive grounded theory.

II. METHODOLOGY

We accompanied a cohort of 94 university introductory astronomy students on two separate visits to a local digital planetarium that had recently been upgraded to a digital system. The mandatory visits formed part of the practical components of the course curriculum. For the first Student Planetarium Visit (SPV1), the course instructor presented a show that they planned with a planetarium operator. For the second Student Planetarium Visit (SPV2), we were given permission to design and administer our own show to the same cohort minus 19 students. In both visits, students were given a questionnaire to fill out during and after the show. The purpose of the questionnaire was to elicit qualitative responses regarding their experiences. SPV1 gave an opportunity to observe and probe the students as they experienced a typical use-case of the facility. SPV2 allowed us to document the role as show designer and operator, while also allowing us to frame the visit and instrument for broader data collection from students, thus furthering the theoretical sampling for our GTM-based study. Care was also put into refining the questionnaire administration from our observations such as providing clipboards due to lack of writing surfaces and displaying questions on the dome due to the dim lighting.

According to an end-of-semester course survey, about 60% of the student cohort had not visited a planetarium prior to this course. The visit to the planetarium, especially as a formal learning environment, could thus be assumed as a relatively new experience for most of the students.

A. Show Design

SPV1 was planned and administered by the course instructor with the help of a planetarium operator. Instructional goals included visualizing the night sky and celestial sphere in “3D”, showing a few of the main constellations and how they indicate cardinal directions, showing earth’s precession while debunking astrology, showing how the Sun “moves” at noon over the course of the year, and convincing students that the planetarium was a great place to go. The show lasted approximately 35 minutes while the students were given 10 and 15 minutes for the pre- and post-show questionnaires respectively.

For SPV2, we designed a show to incorporate aspects unique to the digital planetarium that were not used in the first show. For example, the content of SPV1 was primarily from a geocentric (earth-based) viewpoint, so we instead made use of an allocentric (non-earth-based) viewpoint[4]. Whereas we decided to broaden the demonstrated planetarium capabilities in SPV2, for the presented content we also planned a narrowed instructional intent as a counter to the broad intent range of SPV1. With these conditions in place, we decided on implementing a Powers of Ten-style[5] astronomical distance demonstration that incorporated the celestial objects covered in the course curriculum. This was inspired by the planetarium study that was conducted by Yu et al. [6] in which the digital dome showed learning advantages over a traditional classroom environment. The distance demonstration was given in two difference segments. In the first, the virtual camera moved away from the earth perpendicular to the path connecting the objects of interest with exponentially increasing circles sizes to demonstrate scale. In the second, the virtual camera moved away from the earth parallel to the path connecting the objects of interest with a numerical counter reading off the distances. The two segments took about 10 minutes each while the students were given 5 minutes for the first part of the questionnaire and 7 minutes for the second part.
TABLE I. Results of the Likert-scale responses of SPV1. Cells shaded green for "ideal" show responses.

<table>
<thead>
<tr>
<th>Rating</th>
<th>Enjoyable</th>
<th>Length</th>
<th>Difficulty</th>
<th>Clarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Not enjoyable</td>
<td>Too short</td>
<td>Very easy</td>
<td>Very confusing</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>13%</td>
<td>45%</td>
<td>1%</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>38%</td>
<td>33%</td>
<td>3%</td>
</tr>
<tr>
<td>4</td>
<td>3%</td>
<td>43%</td>
<td>13%</td>
<td>12%</td>
</tr>
<tr>
<td>5</td>
<td>15%</td>
<td>6%</td>
<td>8%</td>
<td>26%</td>
</tr>
<tr>
<td>82%</td>
<td>Very enjoyable</td>
<td>Too long</td>
<td>Very difficult</td>
<td>Extremely clear</td>
</tr>
</tbody>
</table>

TABLE II. Key ideas as percentage of total ideas that emerged from the analysis of what students indicated was the Best Part of SPV1.

<table>
<thead>
<tr>
<th>Positivity</th>
<th>Segment 1</th>
<th>Segment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Ideas</td>
<td>80%</td>
<td>65%</td>
</tr>
<tr>
<td>Negative Ideas</td>
<td>20%</td>
<td>35%</td>
</tr>
<tr>
<td>Students enjoyed</td>
<td>80%</td>
<td>61%</td>
</tr>
<tr>
<td>Students did not enjoy</td>
<td>20%</td>
<td>39%</td>
</tr>
</tbody>
</table>

TABLE III. Combined table of ideas from SPV2 identified as positive and negative from students along with proportion of students that indicated whether they enjoyed or did not enjoy the segment.

For the qualitative responses, open-coding was initially done at a low-level with subsets of the responses. Overlapping ideas among these codes were grouped into categories, from which high-level key ideas emerged. Further responses were then coded as high-level concepts emerged, ultimately producing a localized model, or grounded theory.

The best-part qualitative question of SPV1 revealed diverging key ideas between an experience of learning (related to teaching material of the show) and one of novelty (related to emotional/affective thought or amusement). The majority fell behind the former (see Table II).

In SPV2, with the questionnaire administered throughout the show, segments could be individually assessed for a higher-grained temporal look at the student experience. Furthermore, the ability to elicit negative responses, and the splitting of the show into separately-probed segments, revealed from the students an evolving critical view of the experience. Qualitative responses were once again coded, this time for references to positive or negative aspects and for other emerging ideas. For the enjoyment rating along with positive and negative ideas, there was an increase in negative experience, revealing an overall decrease in engagement (see Table III).

The other emergent key ideas revealed themes that closely reflected the key ideas from the best-part analysis of SPV1, but with a split into positive and negative versions of learning and novelty. Among student ideas, there was a drastic decrease in positive ideas of the learning experience while both positive and negative ideas of the novelty experience doubled (See Table IV).

III. ANALYSIS AND RESULTS

Results were digitized through scanning in student responses. The resulting PDFs were loaded into the NVivo software where individual responses could be sectioned and coded.

A. Student Enjoyment

The Likert-scale question of SPV1 and indication of enjoyment were tallied. The results in Table I revealed that the SPV1 show was highly engaging.

B. Engagement with Educational Content

In assessing how students engaged with the educational content, we first coded for key ideas from what students claimed they learned from the show in SPV1. For SPV2, we coded what students said each segment was about. The emergent high-level key ideas could be compared with what the identified "instructional intent" was from the observational data. This included what the instructor of SPV1 said their
goals were for visiting and for SPV2 it was our own recorded goal. The results revealed high levels of engagement with the educational content in both visits, though in SPV2 there was a slight decrease between segments (see Tables V and VI).

Content reflection questions were also given systematic ratings of quality for a further assessment on engagement with educational material. For SPV1, this was done for how close the sketch of the celestial sphere was to a textbook case. For SPV2, this was done for how close their description reflected the instructional intent. Students showed very little change from Pre-show to Post-Show in SPV1 and there was more significant disengagement in SPV2 between the two similar show segments (see Table VII).

### C. Localized Model of Student Engagement

The similar emergent categories of enjoyment in SPV2 gave us confidence in the themes while the splitting into positive and negative versions gave us further insight into student engagement. In SPV2, we could observe a show become less engaging overall as it progressed with an increase in negative ideas. This was mostly due to the loss of positive ideas associated with the learning content. Meanwhile, the doubling of ideas associated with novelty suggested less engagement with

<table>
<thead>
<tr>
<th>Segment 1</th>
<th>Segment 2</th>
<th>Broad Category of ideas corresponding to Instructional Intent</th>
</tr>
</thead>
<tbody>
<tr>
<td>99%</td>
<td>~89%</td>
<td>Distance</td>
</tr>
<tr>
<td>1%</td>
<td>11%</td>
<td>Not corresponding to specific intent</td>
</tr>
</tbody>
</table>

TABLE VI. Key ideas that emerged from the analysis of how students explained what the individual SPV2 show segments were about.

the educational material, a notion backed by the decreased quality of engagement in the educational content reflection. Furthermore, when asked what the shows were about, a varied instructional intent was reflected by an equally varied recollection by the students.

In considering SPV2, a clear deviation occurred over the course of the visit among the students from a primarily positive learning experience to either a negative experience of boredom or a positive experience of entertainment. In both cases, the ideal case of student engagement with the relevant educational material was lost. To capture this deviation from the positive learning outcome, we identified two semi-abstract, quantifiable domains of the planetarium visit in the context of a student being brought to the digital planetarium to learn. The first domain is how much relevant engagement or sense-making is taking place for the student. It was necessary to explicitly narrow down this domain because the student could leave the planetarium having "learned" something that was unintended by the instructor. We therefore considered this domain to be how much material relevant to the instructional intent the student is actively attending to and therefore engaging with. The other domain is attentiveness. This domain captures both how many things the student is aware of along with the degree to which they are attending to those things. We modeled here the experience as a two-dimensional spectrum regarding the two variables of attentiveness and relevant engagement (see Fig. 1). This we referred to as our localized model of student engagement, which addressed GQ1.
IV. CONCLUSIONS

The broad purpose of the Grounded Theory Method is to obtain insight into a complex situation by producing a localized model or "theory" that is grounded in the data at hand. Thus, the outcome of the present phase of the work is the localized model shown in Fig. 1 that helps to assist in designing meaningful learning experiences. There are indeed unique affordances of the digital dome that must convince instructors to bring their students to the facility. These affordances should provide net benefit to the curriculum at large to make the visit worthwhile. Our localized model suggests there is an optimal level of attentiveness in a digital planetarium experience that provides maximal relevant engagement with the learning material provided by the instructor. It also indicates that attentiveness can be enhanced or can be compromised by elements of the experience, i.e., that students engage with a digital planetarium show through a combination of both the learning and novelty elements. If there are not enough of any of these elements, then boredom can set in and students may not engage with any part of the show, including the relevant learning material. On the other hand when there are too many elements, such as entertaining physical features of the immersive dome or a large collection of expected learning outcomes by the instructor, then the engagement with the relevant learning material can similarly be lost.

It should be emphasized that the present paper only presents a model based on the student perspectives. In a future paper, we will show how these findings were combined with another model that resulted from the simultaneous GTM analysis of observational data to address the investigation's GQ2. For this synthesis, we introduced the GTM notion of a "theoretical supplement", specifically with Working Memory[7] and Cognitive Load Theory[8], in order to further model how these aspects can first be influenced by the instructor in the planning phases and then later compete for student attentiveness in the manifested instruction. This combination of the two localized models we named the Model for Curriculum Design in the Digital Planetarium (MCDiP).

As with any grounded theory, its quality is only as good as the amount of theoretical sampling that accompanies its construction. In order to form a more robust model, more studies need to be done to introduce data that captures the wide diversity of experiences that digital planetariums have potentially introduced. This includes aspects stemming not only from the different types and styles of the facility, but from the different use-cases as well. This would provide further tests of our model while giving opportunity for further refinement.

A. Acknowledgements

We would like to thank the UCT PhAsER group, IDIA, and Iziko Museums for their assistance with the investigation.

V. APPENDIX

A. SPV1 Questionnaire

The following was administered to students after the planetarium show of SPV1. Q3 and Q4 were also administered prior to the show.

Q1) Your friend comes up to you after the planetarium show. He says, "I missed the show! What was the best part of the show?" What do you tell your friend?

Q2) Your friend then asks, "So, what did you LEARN from the planetarium show?" What do you tell your friend?

Q3) Your friend then says, "Can you explain to me what you know about the celestial sphere?" Explain to your friend what is meant by the celestial sphere and why it is useful.

Q4) Your friend then says, "Can you draw me a quick sketch of the celestial sphere?" Make a sketch for your friend in the space below.

Q5) Pick from 1 to 5.
   a) The planetarium show was:
      Not enjoyable (1) Very enjoyable (5)
   b) The show was:
      Too short (1) Too long (5)
   c) The information presented was:
      Very easy to understand (1) Very difficult to understand (5)
   d) The show was:
      Very confusing (1) Extremely clear (5)

B. Questionnaire SPV2

The following was administered to students during the planetarium show of SPV2. Q1 and Q2 were administered after both show segments. Q3 was administered only after the second segment.

Q1) Student A says "I enjoyed this show very very much!"
Student B says "I did not enjoy it that much."
With whom do you agree? Explain your choice in detail.

Q2) Your friend missed the show and asks you what it was about. Write down what you would tell your friend.

Q3) Student A says "I learned more from the first half of the show."
Student B says "I learned more from the second half of the show."
With whom do you agree? Explain your choice in detail.
Unlearning indoctrination: Tensions between decolonizing curricula and characteristics of whiteness

Mathilda J. Smith¹ (she/her)
¹Lyman Briggs, Michigan State University, 426 Auditorium Rd, East Lansing, MI, 48823

Michelle N. Brown² (she/her), Abigail R. Daane¹ (she/her), and Clausell Mathis¹ (he/him)
²Curriculum and Instruction, Pennsylvania State University, 144 Chambers Building, University Park, PA 16802
³Department of Physics, South Seattle College, 6000 16th Ave. SW, Seattle, WA, 98106

Understanding how to make physics instruction more equitable is an ongoing challenge. While many teachers desire this change, few resources support teachers in actualizing DEI efforts in classrooms. In this study, we observed a learning community of physics instructors attempting to decolonize curricula, i.e., critically analyzing practices and content within instruction that decenter the Eurocentric physics narrative. We use a critical whiteness analytical framework to examine tensions instructors encounter as they attempt to reform traditional curricula together. Findings show four characteristics of whiteness and their antidotes across the data: 1) Perfectionism, 2) Sense of urgency, 3) Quantity over quality, and 4) One right way. We found the democratic structure of the group restrained characteristics of whiteness as teachers worked to unlearn norms of schooling. By attempting new ways of thinking to decolonize the curricula and meetings, they productively moved towards a middle ground between the characteristics, and achieve their antidotes.

Keywords: decolonizing curriculum, tensions, characteristics of whiteness
I. INTRODUCTION

Awareness of “decolonizing” curricula within educational research has increased within the last decade [1,2], and is an important way to include a wider range of students’ worldviews, cultures, and identities into schooling [3]. Although there are multiple definitions to describe how to “decolonize” curriculum, all focus on recognizing and interrogating the role of Eurocentrism in institutions today [2]. Our definition of decolonizing curriculum aligns with a process of “(a) recognizing constraints, (b) disrupting, and (c) making room for alternatives” [1]. This entails critically examining how knowledge is portrayed in the physics curriculum, questioning what and/or how work is excluded, and intentionally including cultural examples and perspectives outside of the white, Eurocentric dominant lens. Despite much discussion of decolonizing the curriculum, there are few examples of what this looks like in practice, particularly in STEM disciplines such as physics, and in virtual spaces [1,4]. Our study seeks to add to this literature by sharing the tensions that emerge as a group of physics educators engage in a virtual community to decolonize their physics curriculum. We enlist a critical framework to ensure we consider the ways even a decolonizing group perpetuates inequities. We also aim to describe how tensions were related to the task of decolonization, and how this group actively resisted the colonial framings of learning communities.

II. ANALYTICAL FRAMEWORK

Our study adopts Critical Whiteness Studies (CWS) [5] as a lens to understand how a group of physics teachers attempt to decolonize their curricula, and the tensions that arise in this process. We uplift second-wave CWS [6] which recognizes whiteness as an ideology. We move away from thinking about whiteness as a social construction of race tied to privilege (i.e., focusing on white identity), and instead consider how whiteness shapes power structures and social hierarchies in often hidden ways. We define CWS as “a framework to deconstruct the material, physical, emotional, and political power of whiteness” [5] Our paper centers Okun’s [7] “characteristics of whiteness” as the main tool to interrogate how whiteness both persists and is countered as teachers seek to decolonize their curriculum. Although Okun’s characteristics were not created based on empirical research, numerous scholars have substantiated these attributes [8,9,10].

Tema Okun and colleagues have authored ongoing work to characterize whiteness and how it operates. Aligned with CWS, Okun positions characteristics of whiteness and white supremacy culture as hidden norms that are ubiquitous, regardless of one’s race. Thus, we expect these characteristics to show up in any community, including one that seeks to decolonize curricula. Okun [7] identifies fourteen characteristics of whiteness, along with “antidotes” which oppose them. We will be focusing on four of these: perfectionism, a sense of urgency, quantity over quality, and only one right way. “Perfectionism” centers inadequacy, and internalizes mistake-making as a personal fault. “Sense of urgency” prioritizes timeliness and a need to reach decisions quickly, and “quantity over quality” values measurable outcomes over process. Lastly, “one right way” assumes an intrinsically correct way to operate. Similar to Okun’s antidotes, we also recognize ways in which physics teachers in our study moved away from these characteristics.

II. METHODS

We aim to share what decolonizing physics curricula can look like in practice. Specifically, we seek to answer the following research questions: What are the tensions that arise among nine secondary physics teachers attempting to decolonize a physics curriculum? How can teachers navigate obstacles to decolonizing physics curriculum?

A. Data collection

Eight physics teachers who attended a conference designed to help physics teachers incorporate social justice pedagogy decided to start an online, voluntary group to discuss various ways to decolonize physics curricula. At these monthly meetings, teachers examined curricula, designed lessons, and critiqued both pedagogy and knowledge in physics teaching practices. Meetings were intentionally structured to provide as much equality in the decision making and power of each member by rotating roles and making decisions through agreement, herein referred to as “democratic”. Two researchers joined the group and began to video-record the meetings via teleconferencing online (zoom). The videos were collected over a span of eight months. Each video was ~1hr and 30 min. The participants were given pseudonyms: Ruby, Layla, Khary, Gwen, Sharon, Bill, Pearl, Yolanda, Cameron, and Jacqueline. All participants were either physics instructors at the secondary, two-year college, or university levels, or physics education consultants. All participants are located in various parts of the United States and Europe.

B. Data analysis

Transcripts of teacher videos were created and analyzed using Vosaic, an online video analysis software. Analyses of transcripts were conducted using an open coding analysis [11] where common themes of tensions between the teachers the first and second author identified and organized themes in an iterative process. Coding was done over two rounds. The first round of coding was identifying tensions and listing them in a document. In the second iteration, we defined parent codes as major tensions. In our final
iteration, we classified the major tensions into three themes of characteristics of whiteness and their antidotes. These findings were shared with the members of the group. Through this process of member-checking [12], small adjustments were made to the descriptions of the themes to align their perceptions of the meeting tensions with the researchers’ themes.

III. FINDINGS

From the data, we found that teachers negotiated tensions between characteristics of whiteness and their antidotes across three themes: perfectionism and urgency vs. process; one right way vs. many paths forward; and quantity vs. quality.

A. Perfectionism/urgency vs. process

Throughout the interview transcripts, we found examples where teachers undulated between a sense of urgency to develop products and a recognition that they need to move slowly and carefully to do the work of decolonizing curricula. During the meetings, there were several moments of tension between preparing and deliberating about what it means to decolonize physics curricula versus actually attempting to do it with a more hands-on approach (e.g., critiquing a lesson). As a result, there was a hesitancy to get started, and frustration at this hesitancy. This is reflective of the push and pull between perfectionism and a sense of urgency versus a process that requires time to learn and make mistakes.

Several participants expressed frustration that meetings were getting too caught up in details and going in circles. For example, in a meeting regarding lesson planning, Khary spoke about feeling that perfectionism was holding the group back from making progress:

I think we’re getting into the place of perfect being the enemy of good. I just don’t know what the hang up is. Let’s just get our hands dirty already [...] I feel like we’ve gotta just take something and workshop it and then move on. We should be on something else... But we are stuck here, and I want us to move forward, however that looks. So if it means taking what’s already been created, and working with that, then let’s do it.

This quote recognizes one characteristic and exemplifies another. Rather than finding a place of compromise, members of the group felt that in order to counteract the negative effects of one characteristic, perfectionism, they must then go in the opposite direction - a sense of urgency. Other teachers similarly expressed the feeling that while what the group had accomplished in previous meetings was important, there was a sense of urgency and a need to stop going over the same ideas and move on to the next step.

Whilst participants were ruminating over why progress seems to have slowed, Sharon articulated in the following quote the group’s sense of being stuck, having completed the ‘easy’ tasks and failing to progress because the tasks are becoming more daunting/difficult:

It could be that we’ve run out of low hanging fruit and now we’re just doing harder work, and therefore it’s slower. It could be that. But I personally feel like the motion has slowed down and... I like moving fast. So when I feel like things are slow, I’m like, ‘oh, therefore problem,’ right? But that doesn’t mean ‘therefore a problem,’ that means ‘therefore slower.’ And so I’m working through, kind of, these intricacies in my mind.

In this quote, Sharon points out that perhaps things seemed to go faster in the beginning because there were ‘low hanging fruit’ to work on first. Sharon exhibits an awareness of how, up to this point, perfectionism hasn’t slowed momentum because the tasks that the group had completed up to this point have not been relatively difficult. However, now that the goals have become more challenging, she challenges herself to not view the slower progress as a problem. Rather, it is indicative that perhaps the group should become accustomed to a new pace, or a new method of working together. Thus, Sharon exemplifies a compromise between sense of urgency and perfectionism - leading to a better appreciation of the process.

B. One right way vs. many paths forward

Another characteristic of whiteness that presented itself in the group, one right way, revealed tensions that emerged while developing and enacting curricula. For example, two participants, Yolanda and Khary, have a conversation about Yolanda’s struggles with the idea of co-creating language with students as it wasn’t the way she was taught:

Yolanda: I think it’s like my own indoctrination. I just think that I was not taught that way. I was taught, you know, the Professor was the holder of knowledge. These are the laws, these are whose laws... I think it’s touching on my own education, actually. And just making me uncomfortable and almost like ‘if I’m doing this in the classroom, like, am I failing as a teacher?’

That sounds so weird. I don’t know, this turned into a therapy [session] but... there’s just something so different about it than I’d ever learned physics, and even taught it. Like, you know, very much I taught it in the past of ‘this is the canon.’

Khary: It feels almost like it’s scary because it’s not the way we’ve learned or even taught ourselves. But we have to do it. I feel like the fear is a sign that it’s like a thing that needs to be done, right?

In this interaction, we see that Yolanda describes her feelings of discomfort and tension with pushing back against the norms of colonized physics curricula due to her own indoctrination. Khary then replies that these feelings could be an indication that the group is moving in the right direction as there can be comfort in the constraints of ‘one right way.’ In contrast, navigating ‘many ways forward’ can be daunting. In a similar vein, when discussing how to
decolonize force lessons, Cameron questions if it would be acceptable to present a previously canon topic in physics curricula as a problem to students:

As far as why this is a problem, why the story of Newton is a problem... For them to answer that question, them being the students, for them to answer that question they would have to believe it is a problem, right? Or we would have to tell them that it's a problem. Is that okay for us to do?

As a result of indoctrination, members of the group struggle with battling against ‘one right way.’ As seen in this quote, a new qualm is introduced in not only if they should push against the canon of colonized physics curricula, but how to do so in an appropriate manner. Comparatively, when considering how to design the curricula and what topics to discuss, Sharon wondered how much divergence this group can make from the canon. This is because she recognizes that students who must be assessed on physics outside of the classroom setting may be expected to know certain concepts in a certain manner. Sharon questions how much the group may be over optimistically designing their curriculum, which portrays a worry that the decolonized curricula may not serve students appropriately if they may later be assessed based on colonized expectations of physics requirements.

C. Quantity vs. quality

The final characteristic of whiteness that was negotiated in the data was the concept of ‘quantity over quality,’ where an organization measures success by how much is produced. In the group, however, participants are aware that prioritizing quantity may have negative effects on the quality of the curricula. For example, when discussing what members expect students to ‘know’ by the end of a unit and what students should be assessed on, Pearl expressed concerns that the quantity may be too large for her classroom:

I’m imagining trying to do this with my ninth graders, and it feels like there’s a lot of pieces to it. And it’s almost like there’s so many pieces, I don’t know if any one of them will get done well. And so I’m curious if our ideal situation would be these mega tasks that include lots of things, or if they’d be more like a few smaller tasks that get at more targeted things we’re trying to see?

This quote exemplifies a tension that the group may be trying to implement too many components into assessments and take on too many tasks. Pearl is thus aware that too many ‘pieces’ may end up creating a product - referring to a completed lesson plan and/or completed curricula - that is not ‘done well,’ and asks the group if this is the ‘ideal situation.’ In another meeting, Pearl brought up the point that prioritizing the ‘big picture’ can eventually lead to a design that is not feasible for teachers. Therefore, she argued that it is important to start at a smaller level:

I also think that we continue to talk in these really big pictures - like ‘and this, and this, and this,’ and I think if we just force ourselves to actually make it work for a context, then we could expand from there, and modulate from there. And all the modulation can happen. But it’s easier to modulate once you have something that you’ve done and [we can] see what worked and what didn’t work. I feel like if we start with a modulation, my personal bias is that I think we might run the risk of making things that are... not feasible or feel like there’s like lots of questions about how they actually work.

By getting down a rough outline of the goal, and then ‘modulating’ it afterwards to better suit their needs, Pearl argued that it is less likely the end product will be unfeasible. However, by prioritizing modulation (quantity), she worried that the group runs the risk of making an unfeasible curriculum that doesn’t actually work when implemented. Ruby also contributed to the conversation of quantity over quality in an earlier meeting when she suggested that the feeling of ‘stuckness’ may be due to the fear of starting the process of creating a ‘product.’ She then referred to a potential method of how to move forward past this fear by taking an already made lesson plan and ‘playing’ with it until the group was comfortable with the product - thus prioritizing quality.

IV. DISCUSSION

Our study makes space to better understand how teachers can recognize and disrupt colonized curricula and make space for alternatives, which is much-needed in the literature [13]. Using a CWS frame allowed us to identify how even within a decolonizing space, tensions tied to whiteness emerged along three areas: perfectionism/urgency, one right way, and quantity over quality. Each of these tensions created opportunities to push on the normative structures of learning communities. From reviewing our findings, we found that this group navigated these obstacles through 1) their articulation of the tensions, 2) the democratic nature of the group, and 3) their feelings of discomfort while moving toward a middle ground between characteristics of whiteness, and achieve their antidotes as they sought to forge new paths away from a Eurocentric curriculum. Before discussing these ideas, we want to acknowledge that characteristics of whiteness are not a unique finding to this group, nor were they exhibited to an unusual amount. These characteristics permeate into all spaces, even those that are constructed to decolonize whiteness. Members of the group were actively trying to counteract these characteristics, as can be seen in the findings.

Throughout our findings, participants were constantly bargaining between characteristics of whiteness and their antidotes, often working in a middle ground between the two. The duality between the tensions depicts how the group was conscious of enacting whiteness whilst trying to
unlearn the norms of schooling that are tied to whiteness; this creates difficulty when trying to navigate and escape this cyclic pattern. A good example is Khary’s quote regarding ‘perfect being the enemy of good.’ Here she acknowledges the problem of perfectionism, to which the group responded by feeling a sense of urgency to work faster, shifting towards another characteristic of whiteness. This arose from the group attempting to focus on too many details, which detracted from the group’s core values and direction, and hindered the feeling of momentum towards their goals. However, Sharon expressed that though ‘things are slow,’ that did not necessarily indicate a problem despite her tendency to prefer ‘moving fast.’ This middle ground is seen again in the tension between prioritizing the creation of a decolonized product over embracing practices that push against whiteness. For example, Pearl highlighted that having too many ‘pieces’ may result in a product that will not ‘get done well.’ She also posits that there may not be a one-size-fits-all manner to accomplish every task. Both of these examples illustrate how the group moved between characteristics of whiteness and their antidotes.

When considering a reason behind the source of characteristics of whiteness being at odds with both each other and their antidotes, the democratic nature of the group may play a factor. Without a leader dictating what is ‘right,’ there can be difficulties knowing how to move forward when something is in question, creating a hesitancy on how to proceed; hence Cameron’s question regarding how the group should go against the canon of colonized physics curricula. As there were few, if any, examples on how to move in this direction, there was great value in creating their own path i.e. one right way vs many ways forward. Sharon exemplifies this notion when she said:

*One thing that is phenomenal about this group is that we don’t have a single leader and as a result, any assertion I make about what next steps should be, who knows if that’s what the others feel right with? We don’t have any single person who’s deciding where we go. I have never been in a group this large that didn’t have a leader that was so functional and that’s really cool. And so we don’t want to lose that.*

Despite experiencing this struggle, all members of the group valued their democratic system. Though not having an individual voice to have a final say could contribute to the push and pull between characteristics of whiteness and their antidotes across the three themes, holding to this democratic structure ultimately creates an environment where the group can combat the notion of ‘one right way’ in a decolonizing space through the collaboration of different ideas to forge a new path. Thus, they are able to mitigate tensions between the constraints of norms of schooling and decolonization of schooling.

The findings from our study show that unlearning whiteness while, in tandem, seeking to decolonize a physics curriculum is a disruptive act. Members in the group must work through these tensions while steering through unknown territory, hence Khary’s feeling that “it’s scary because it’s not the way we’ve learned.” Hesitancy also was a byproduct demonstrated by Yolanda when she expressed her struggle with diverging from what she was taught in the past, and changing her expectations and frame of thinking. To Yolanda, it felt ‘uncomfortable’ to bifurcate from her indoctrination.

**V. CONCLUSION & IMPLICATIONS**

While there have been several groups of educators working on equitable curricula [14-17], we have shown that this work can come with challenges as educators develop methods that reflect inclusivity. This is likely caused, at least in part, by the lack of resources in actualizing DEI efforts in physics classrooms [18]. In this study, we used a critical whiteness framework to examine tensions that arose whilst teachers attempted to decolonize the Eurocentric and white narrative of physics. Our analysis demonstrated ways in which participants navigated whiteness, particularly through their careful engagement with each other, the democratic nature of the group, and how they managed feelings of discomfort. Teachers had to unlearn the norms of schooling and attempt new ways of thinking to decolonize their curricula - which created a push and pull between characteristics of whiteness and their antidotes.

An implication from this study is that teachers can recognize and disrupt the dominance of Eurocentric physics curriculum by creating carefully-structured professional learning communities (PLC). Future research and professional learning spaces should consider the ways in which the structure and group norms influence how members decolonize curricula. We recommend that the PLC structure aligns with a democratic culture, that discomfort is recognized as a norm of this process, and that there is an acceptance of the fact that characteristics of whiteness can be found even in spaces designed with the intention of interrupting whiteness. With these recommendations, other educators and educational researchers may be able to more effectively navigate and actively resist the colonial framings within their learning communities. We recognize the need for further research that captures the actual processes educators undergo as they attempt to decolonize their curricula. Also of interest is research that considers the ways in which characteristics of whiteness harm and/or benefit educators as they decolonize curricula...In addition, our study has found the importance of using a critical lens to recognize the ways in which whiteness co-exists in all spaces, including those that move toward decolonization.

**ACKNOWLEDGMENTS**

We gratefully acknowledge all the teachers in the group for their generosity in making their meetings and ideas accessible to the team. We are grateful to our respective institutions for supporting this work.


A case in which canonically incorrect ideas do not hinder conceptual progress in introductory physics

Al K. Snow (they/them)
Department of Physics, University of Washington, Box 351560, Seattle, WA, USA, 98195-15603

Paula R.L. Heron and Lauren C. Bauman (she/her)
Department of Physics, University of Washington, Box 351560, Seattle, WA, USA, 98195-15603

Amy D. Robertson (she/her) and Lisa M. Goodhew (she/her)
Department of Physics, Seattle Pacific University, 3307 Third Ave W, Seattle, WA, USA, 98119-1997

Resources theory depicts resources as dynamic, context-dependent pieces of knowledge, and defines learning as building from students’ resources. In this paper, we will use a classroom video example of students working through ACORN (Attending to Conceptual Resources in) Physics tutorials, resources-oriented instructional materials for introductory physics, to illustrate a learning sequence in which one group of students make progress towards a model for what makes a lightbulb light, even as they discuss ideas we consider canonically incorrect. This case serves as an existence proof that canonically incorrect ideas need not hinder conceptual progress, challenging historical models of misconceptions as obstacles to learning.
I. INTRODUCTION: CANONICALLY INCORRECT THINKING IN STUDENT LEARNING

When students enter the classroom, they already have ideas about physics that they have developed from their experiences living in the physical world. Some of these ideas are canonically incorrect—that is, they do not match a physicist’s description of the same phenomena. The misconceptions framework frames many incorrect ideas as misunderstandings that need to be specifically confronted and addressed, because they can be hard to change [1–3]. The misconceptions framework thus informs a number of instructional strategies that elicit and resolve common incorrect ideas, and turns instructor attention toward helping students overcome incorrect ideas.

Resources theory, another theory of cognition and learning, suggests that students’ ideas are the basis for learning, and even unrefined or inaccurately-applied ideas can contribute to new knowledge [4]. "Resources" are pieces of knowledge that students have inferred from lived experience [5]. The activation of resources is context dependent; activation can be appropriate or inappropriate [6]. In the resources framework (and similar theories), learning is seen as the refinement and reorganization of resources, which are diverse and interrelated within a complex system [4]. The resources framework directs instructors attention toward "seeds of knowledge" that can be drawn out/built from during instruction, which may mean allowing incorrect ideas to persist so they can be refined, sometimes in order to foreground other "seeds."

In this paper, we analyze a classroom video excerpt where students use resources, deployed in both canonically incorrect and correct ways, to make progress towards a model for circuits in physics. This case study [7, 8] illustrates that canonically incorrect ideas need not hinder conceptual progress and can be a part of a cascade of ideas that lead to canonically correct models, challenging a model of misconceptions in which every incorrect idea needs to be addressed. Our findings support a resources framing of student ideas, and offer empirical support for instructors who may be concerned about impacts of leaving students with canonically incorrect answers [9].

II. INSTRUCTIONAL CONTEXT: ACORN PHYSICS CIRCUITS TUTORIAL

The video excerpt we examine comes from a university physics class session in which students worked through an ACORN Physics tutorial about circuits. This worksheet is designed to elicit common, potentially-fruitful ideas (conceptual resources) about electric circuits. The literature identifies conceptual resources for understanding circuits such as: voltage drives current, resistance limits current, and the way circuit elements are connected affects current [10, 11]. Because previous research has identified these ideas as somewhat common, we expect that some students will use ideas like these to reason about the electric circuits presented in the worksheet. At the same time, we expect that the particularities and frequency of the activation of these resources may be different in different contexts, given the dynamic, context-sensitive nature of resource activation [4]. The worksheet is sufficiently open-ended so as to encourage the sharing of additional ideas unanticipated by the worksheets.

This tutorial prompts students to sense-make [12] about a set of electric circuits composed of ideal wires, light bulbs, and a battery (see Fig. 1). Many questions in the worksheet present information (e.g. ranking of brightness of bulbs in the circuit) and ask students to explain rather than predict. In the case we discuss here, students are working through a version of the worksheet that has since been modified.

III. CONTEXT & METHODOLOGY

The case being analyzed is from a small-group discussion that took place in a calculus-based introductory physics course at a small (<5000 students), private, liberal arts university in the Pacific Northwest. The course mainly serves students majoring in physics, engineering, computer science, chemistry, and biology. Video data features small groups of students (3-5) during the portion of class dedicated to discussion and problems. The class in which our data was collected was the third of a three-quarter introductory physics sequence, composed of approximately 30 students, taught by one faculty member and supported by two undergraduate Learning Assistants. The racial and gender demographics for the population is as follows: 48% female and 52% male; 7% international students; U.S. resident students are 44% white, 17.8% Asian, 8.3% Black, 0.3% Hawaiian Native/Pacific Islander, 14.9% Hispanic of any race, 8.1% two or more races.

In reviewing the video record, we looked for clips that: (i) were rich in dialogue between students, (ii) featured multiple speakers/turn-taking, (iii) included multiple ideas that evolved throughout the conversation, and/or (iv) included instructional moves from the course instructors that seemed to help students make conceptual progress. The clip we selected for this analysis shows a group of students collaboratively working to make sense of how a lightbulb works, satisfying criteria i, ii, and iii; we selected the clip from a corpus of over 8 hours of classroom footage. Our candidate clips were discussed among a team of 6 researchers, composed of faculty and graduate & undergraduate students.

1 The university reports gender demographics as male and female, but we do not assume this fully represents the spectrum of gender identities held by students in this population.
Our team met over several months to watch and discuss video data collected for this project. We entered the analysis phase with broad thematic questions of how students’ ideas developed during the class period, and used an inductive approach to refine our research question and claim [13]. We iteratively watched and discussed the focal episode for this paper, discussing possible interpretations of and claims from this episode [14]. After several iterations of reviewing the video and discussing possible interpretations, our claim was narrowed to the one we make here: these students make conceptual progress even as they bring forward canonically incorrect ideas. We used case study methods [7, 8] to develop this claim, situating it as a case of learning in the resources framework; this is a case that demonstrates a specific way in which incorrect idea can promote conceptual development. Here, the canonically incorrect idea (that charge gets “used up”) is not itself the thing that gets refined; instead it acts as a catalyst to help students recognize other things they do know.

IV. TRANSCRIPT ANALYSIS

A. Incorrect ideas do not hinder conceptual progress

_Transcript_. This case highlights a discussion among 4 students as they work through the ACORN Physics circuits tutorial in class. Using the PhET Capacitor Lab and Circuit Construction Kit: DC simulations [15] in conjunction with the tutorial, they discuss how charge, energy, and potential difference are related to one another, and how different circuit elements (battery, capacitor, and bulb) function. The PhET Simulation allows them to build and manipulate the circuits in the tutorial, confirming or challenging their ideas as they go. Here we examine an approximately 10-minute excerpt from the students’ conversation about the questions in Fig. 2. The four students in conversation are pseudonymed as Akiko, Ben, Cole, and Daniel, and the instructor is pseudonymed as Zoe. We present a transcript excerpt in this section and examine a particular canonically incorrect idea that emerges, illustrating how it plays a role in a sequence where students make conceptual progress without being explicitly addressed.

The discussion begins with Akiko using the PhET Capacitor Lab simulation to observe that a lightbulb lights up, then dims and goes out, after it is connected to a charged capacitor. Akiko says that they "don't know why it [dimmed] though." Daniel proposes an explanation that links charge flow, light bulb brightness, and capacitor charge:

1. Daniel: I’m thinking it’s because like charges move across and then it’s, once each plate, like, has no more charge then no more charges are moving...So it moves for a little bit and then stops moving. And that’s why the light starts really bright. 'Cause lots are moving and slowly they move less and less and less. And then nothing moves 'cause the two plates are the same. That’s how I would explain it.

2. Akiko: Right.

3. Daniel: There’s no more potential difference, right? (Daniel then notices the camera and has a joking conversation with their tablemates about it.)

4. Ben: ...Um, potential difference become zero. Which means there’s no flow.

5. Daniel: ...Could you repeat that?

6. Ben: Like saying the system loses its potential difference.

7. Cole: Because it’s, like, it’s initially charged...And it, it uses that charge until it uses up the—whatever initial charge you had.2


9. Ben: And then there’s no flow of charge.

In this exchange, Daniel leads by proposing that the bulb shines the brightest when lots of charges are moving, and as the capacitor discharges, less charge moves and the bulb dims (line 1). Akiko affirms this (line 2), then Daniel proposes a second explanation (line 3), bringing in the potential difference of the capacitor, implying that when charges have stopped moving, "there’s no more potential difference." Ben revoices this (line 4), stating explicitly that when there is no potential difference, there is no flow of charge. In line 6, Ben elaborates that the system loses its potential difference as the capacitor discharges. Cole (line 7) adds on that a capacitor supplies charge until it is fully discharged. In this exchange, these students make the connection that in this circuit, potential difference, capacitor charge, current, and light bulb brightness are all related; as one decreases, the others also decrease.

The students next move on to discuss question C of section I: “What makes a light bulb light up?” They create an analogy for current, comparing charge flow to a river.

10. Daniel:...I’m just looking at the next question [question C]. Like what about electrons moving makes something light up? Just, I don’t know.

11. Ben: That’s just how a light bulb works. Gotta figure that out.

12. Daniel: I’ve always been curious about this because like you’ve got a river and a river’s flowing. You put like something into it that gets electricity. So is it like

---

2 In some cases, we’ve removed parts of transcript where students issue affirmatives ("yeah" or "right") or ask for clarification (e.g., "which question is this?") when these utterances are not central to the conversation.
In line 7, Cole explains the dimming of the bulb progress. the filament and this causes the bulb to light (line 17). Instructor, Daniel restates the idea that electrons flow through to questioning if the filament of a bulb causes the bulb to light charge on a capacitor decreases until it runs out (Cole, line 7) resizing that charge flow is directly related to bulb brightness. Although Akiko emphasizes that flow of charges, and seem to make a connection to the way that hydropower harnesses the energy of river flow to create electricity. Then, Ben shifts gears (line 13), bringing up the fact that inside of a light bulb, there is a “wire,” and Daniel adds that it has “high resistance” (line 14). Daniel proposes that the wire “loses” charge to heat and light (a canonically incorrect idea), using similar language as in lines 6 and 7 when Ben and Cole suggest that the system loses potential difference as the capacitor discharges. Akiko (in line 15) takes the idea of “charge loss to heat and light” and brings in energy, specifically referring to energy transformation from one form to another. Daniel then brings multiple ideas together, saying: 16. Daniel: So the wire that’s in the light of the light bulb has really high resistance... electrons go through and it heats up and emits light.

In this 16-line exchange, the idea that charge needs to flow in order for the light bulb to light, originally mentioned in line 10, evolves to become the idea Daniel expresses here: that electrons flow through the filament of a light bulb, causing the filament to heat and therefore illuminate. Although Akiko explicitly mentions thermal energy, Daniel’s thinking in line 16 does not take this idea up immediately.

The instructor joins the table just after Daniel’s statement (line 16) and asks what makes a light bulb light up. Daniel repeats the general model they mentioned in line 11:

17. Daniel: I was thinking that the electrons flow through the wire and if the wire, like the filament, if the filament has really high resistance, then that means it’s like gonna heat up and possibly like be lit.

So far, we have seen the group of students go from hypothesizing that charge flow is directly related to bulb brightness (Daniel, line 1) to recognizing that when potential difference is zero there’s no flow of current (Ben, line 4) to realizing that charge on a capacitor decreases until it runs out (Cole, line 7) to questioning if the filament of a bulb causes the bulb to light (Daniel, line 14) to knowing that energy gets transformed into thermal energy (Akiko, line 15). In summarizing for the instructor, Daniel restates the idea that electrons flow through the filament and this causes the bulb to light (line 17).

A canonically incorrect idea does not hinder conceptual progress. In line 7, Cole explains the dimming of the bulb in terms of the charge getting “used up”; “it [the capacitor] uses that charge until it uses up the—whatever initial charge it had.” In line 14, Daniel uses a similar idea: “Is it a wire that’s really, like, has high resistance? And so it loses that charge to heat and light?” From these two statements, we infer that the group may be thinking that the charges initially stored in the capacitor could be used up to produce light and heat, or transformed (as energy is) into light and heat.

This idea—that charge is “used up” in a circuit—is closely related to the commonly reported misconception “current is used up” [16]. Misconceptions research may propose, then, that an instructor intervene to address Daniel’s incorrect idea in line 14 [17]. Yet in this 17-line exchange, students make conceptual progress even with this idea as part of their collective thinking. In particular, many different ideas appear that relate current (referred to as “charge flow” or “electron flow”), potential difference, and light bulb brightness. When the idea of charge use/loss appears, it is mentioned in relation to potential difference of a capacitor (line 7)—the system loses its potential difference because the plate “uses that charge until it uses” whatever “initial charge you had”—and the heat/light from the bulb (line 14)—the wire “has high resistance” and so “loses [its] charge to heat and light.”

From the ideas in lines 1, 4, 7 (1: current is related to brightness, 4: zero potential difference means no current, 7: capacitor charge is “used” and decreases until it runs out), we gather that the students have made the following connections: (a) as potential difference of a capacitor decreases, capacitor charge decreases, and (b) when capacitor charge decreases, current and bulb brightness also decrease. We interpret the students as demonstrating an understanding that a flow of charge (from the charged capacitor) is necessary for a bulb to light, which is the idea stated by Daniel (line 16, 17). In this way, this “charge gets lost/used up” idea does not appear to hinder the learning process of the students, and instead can be seen as a catalyst that helps the students figure out what they understand about this circuit scenario.

B. Incorrect ideas plausibly help conceptual progress

Transcript. In this section, we pick up just after Daniel summarizes the group’s progress (line 17). Zoe asks the group a confirmatory question, which Daniel confirms, and then poses a question:

18. Daniel: But since energy is conserved, does that mean it’s losing, does energy rely on how fast it’s moving through the wire?

19. Zoe: Let’s think about this. When there’s two capacitor plates, they have some charges on them. And then when we connect the light bulb, then it lights up and we see evidence of like, the thermal or light energy, whatever you wanna call it. But where was the energy before that? What was storing it before?

20. Daniel: The potential difference?


22. Zoe: So the energy is stored in the interaction between the two plates?


Daniel’s question for Zoe (line 18) references Akiko’s energy transformation idea (line 15). In line 19, we see Zoe
draw students’ attention to the original setup (two capacitors with charges on them) and then revoice Daniel’s narration of what happens when charges flow through the wire (the filament "heats up" and is going to "be like lit"), connecting this to thermal and light energy. Zoe then asks where that energy comes from, connecting Daniel’s question about energy loss to the principle of energy conservation. Cole and Daniel respond that the energy comes from the potential difference (line 20) and the capacitor plates (line 21), implying that before the light bulb lights, the energy must be stored in or between the plates of the capacitor (line 22, 23). The interaction continues with Zoe guiding the students to recognize where the energy from the charged capacitor goes:

24. Zoe: And would it make sense to say that as the light bulb is lighting up and there’s heat being transferred to the environment, whatever like, going out in that direction—is that energy between the plates decreasing?
25. Cole: Yes. Because it’s lost with the environment.
26. Zoe: And if we’re thinking about like, what would tell us how much energy is stored between plates? Like would that depend on?
27. Ben: How bright the light is?

In line 24, Zoe builds on Cole and Daniel’s response that the energy comes from the potential difference/capacitor plates, asking if the natural conclusion, then, is that when the light bulb is lit and loses energy to the environment, the energy between the capacitor plates is decreasing. Cole affirms this, saying that the energy decreases “because it’s lost with the environment” (line 25). In line 26, Zoe poses a question tied to another implication of the students’ proposal that the energy transformed to heat and light comes from the capacitor plate, asking students to think about what qualitative observables would tell them about the amount of energy stored in the capacitor. Ben makes the connection between the brightness of the bulb and the amount of energy in the capacitor line 27, and Zoe affirms Ben’s answer.

A canonically incorrect idea can be productive to the learning process. From Daniel’s question (line 18), we know the students (or at least Daniel) have questions about what is conserved and what is lost, as is common when students are reasoning about circuits [16]. The students do not state with certainty the source of the energy in the bulb; Zoe seems to play a key role in guiding the students to understand where the energy comes from (line 24, 26).

Earlier in the discussion, the following ideas come up: (a) resistance causes a loss of charge to heat and light (line 14), (b) some kind of energy is transformed into thermal energy (line 15), and (c) electrons go through the filament, causing it to heat and light (line 16). With instructor discussion, some of these ideas (lines 14, 15) are combined and reworked into the following idea: the resistance in the filament causes some kind of energy, initially stored in the capacitor (lines 19, 20, 21), to be converted into thermal/light energy (lines 24, 25).

The original, incorrect “charge loss” idea appears to be born out of the notion of conservation, as is the idea that energy from the capacitor is converted to heat and light energy in the bulb. Daniel (in line 14) seems to be sense-making about the question of “what makes a bulb light up,” connecting their first response to the question (in line 10)—electrons are moving—to the analogy they constructed in line 12—something gets put into the flow—to Ben’s proposal in line 13 that lighting up has to do with the wire in the bulb. That is, they connect the resistance of the wire to the emission of heat and light, and ask whether that heat/light comes from charge being lost. They seem to know that something is conserved, but express their attribution of that something to charge as a question. With the additional discussion with Zoe, the students further develop their model; their understanding grows from Daniel’s initial hypothesis that charge from the capacitor is "lost" to heat and light, and becomes the more complex idea that energy from the capacitor is lost to heat and light in the bulb as it discharges (lines 25, 26, 27).

The entire instructor conversation (lines 19-28) is brought on by Daniel’s question (line 18), born out of the lack of clarity the students have with what is being converted/lost and where energy comes from. In this sense, the canonically incorrect idea that charge is "lost" to heat/light plausibly helps students conceptually progress by acting as a catalyst for students to recognize what they do and do not know. The initial charge idea serves as a productive seed that grows into the idea of energy conservation between the capacitor and bulb.

V. DISCUSSION & CONCLUSIONS

In the analyzed excerpt, we see the students go back and forth between using two different circuit quantities—potential difference and charge flow—to explain how a light bulb lights. The charge gets “used up” idea acts as the bridge between these ideas. This incorrect idea did not need to be directly confronted in order for the students to be able to make progress in understanding what makes a light bulb light (for example, an understanding that charge flow is necessary for a bulb to light). Even with this incorrect idea in play, the students still have a fruitful learning experience that ends with several canonical understandings about circuits. This case illustrates a situation where (1) an incorrect idea didn’t need to be confronted overtly, and (2) an incorrect idea plausibly contributed to a productive learning experience.

This excerpt validates what many instructors experience, and what theory claims—that it’s possible for a misconception to (a) not hinder a student’s learning experience, and (b) be productive—despite being characterized as an obstacle [4]. While instructors may feel the need to correct students’ misunderstandings, we see here that making corrections is not always necessary, and that letting students think and build their ideas, even if incorrect, can be a fruitful learning experience.

ACKNOWLEDGEMENTS

This work was supported in part by the National Science Foundation under grants 1914603 & 1914572. The authors thank those who helped with data collection and review.
Conceptual Challenges of Discretizing Wave Functions: A Case Study

Christian D. Solorio, Elizabeth Gire
Department of Physics, Oregon State University, Corvallis, Oregon 97331, USA

To understand the conceptual challenges of discretization that students face, we present an instrumental case study of one student. This student took a junior-level quantum mechanics course and the accompanying computational lab course in the Winter 2021 term. The following year, she became an undergraduate teaching assistant for the computational lab course. Using a video elicitation interview, the participant reviewed a video clip of herself and a partner working on a kinetic energy operator computational activity. During the interview, she reflected on her understanding of discretization and identified two challenges associated with it: recognizing functions as column vectors and interpreting $\Delta x$. These challenges were productive for students in considering the nuances of discretization.
I. INTRODUCTION

The traditional position-first approach to teaching quantum mechanics uses the Schrödinger equation early on to introduce the position basis and the energy basis for the infinite square well (ISW), also known as the particle in a box. In the spins-first approach, the postulates of quantum mechanics are introduced in the context of sequential Stern-Gerlach experiments of spin-1/2 particles, which have a finite, discrete basis. The spins-first approach to teaching quantum mechanics has been shown to improve aspects of students’ learning compared to the traditional position-first approach [1].

While there is good motivation for using the spins-first approach, instructors at Oregon State University have noticed students’ difficulty in transitioning from discrete systems (like spin-1/2 particles) to continuous ones (like the ISW). Computation can be a powerful tool for facilitating this transition, as oftentimes continuous objects (like wave functions) and operations (like integration) must be discretized to be implemented into code. We developed a set of computational activities that introduced some of the ways continuous objects and operations can be approximated discretely. We believe that ideas of discretization can build a foundation for students and help them identify the connections between discrete and continuous systems. To explore this, we ask the following:

Research Question: What are the conceptual challenges of discretization that students engage with?

We answer this question by presenting an instrumental case study of a student, Mary (a pseudonym), who experienced a discretization-focused computational activity as both a student and an undergraduate teaching assistant (TA). Our qualitative analysis highlights some of the challenges of discretization students encounter and the productive elements of those challenges. This case study will inform the ways that instructors can support students’ productive ideas of discretization, which may have implications for students’ building connections between discrete and continuous quantum systems.

A. Instructional Context

Upper-division physics students at Oregon State University take an intensive five-week long, seven-contact-hour quantum mechanics course called Paradigms in Physics: Quantum Fundamentals. This course uses a spins-first approach to introducing quantum mechanics and the course ends with the ISW. In the Winter 2021 term, students concurrently took a computational physics lab course where they used numerical methods to solve quantum mechanics problems. Due to the ongoing COVID-19 pandemic during the Winter 2021 term, both of these courses were taught over Zoom.

B. Broader Research Context

In the Winter 2021 term, Mary was longitudinally observed by one of the authors in the computational lab course. One year later in the Winter 2022 term, Mary became a TA for the computational lab course. The rapport developed between Mary and the observing author perhaps helped her feel more comfortable sharing her thoughts during the interview. Additionally, the rapport was helpful during the analysis process because the authors had the additional context of Mary’s experiences as a student.

II. METHODS

A. Video Elicitation Interview

In the Winter 2022 term, we conducted an in-person, one-hour-long, semi-structured interview with Mary. The interview employed video elicitation [2], also known as stimulated recall [3–5] where a participant views a video of themselves during an event. This method typically results in participants recalling their understandings and feelings and reflecting on their thoughts and actions during an event [2].

During the interview, Mary watched a 10-minute video clip of herself as a student working with a partner, Dave (a pseudonym), on a computational activity (described later in this section). Following the interview protocol, the video was paused at particular time stamps, and Mary was asked to reflect on the portion of the clip she saw, compare her understanding as a student and as a TA, and describe how she would assist students who had questions about the activity.

This interview differs from other video elicitation interviews in that the events of the video clip and the interview take place a year apart. Lyle suggests video elicitation interviews happen soon after the recorded event so that the study participant can better recall their thoughts and experiences [3]. At a few points during the interview, Mary mentioned not remembering what she thought in the moment as a student in the video clip. Because of this, the perspectives she gave during the interview were likely more informed by her experience as a teaching assistant rather than as a student.

Data captured during the interview included video and audio recordings of Mary, video of a nearby whiteboard that Mary wrote on during the interview, and a screen recording of the video clip. The interview was transcribed and reviewed to identify salient experiences, understandings, and perspectives that Mary shared related to our research question.

1. Activity Description

The computational lab course consisted of four activities in the quantum mechanics context developed by the instructor and one of the authors. Each activity of the computational lab course was related to the ISW. The third activity of the
computational lab course, taking place before students were introduced to wave functions in the Quantum Fundamentals course, focused on finding the eigenvalues and eigenvectors of the finite-difference approximation of the kinetic energy operator [6]. The goal of the activity was to demonstrate how a differential operator can be implemented computationally using a finite difference approximation and how wave functions can be discretized.

Students were tasked with first coding a matrix to represent the finite difference approximation of the kinetic energy operator $\hat{T}$. Students were given the equation below to show how the approximated operator acts on a wave function. The equation below shows the finite difference approximation of the kinetic energy operator $\hat{T}$ matrix acting on a wave function, represented by a collection of coefficients $\psi(n\Delta x)$ in a column vector.

\[
\begin{pmatrix}
\hat{T}\psi(\Delta x) \\
\hat{T}\psi(2\Delta x) \\
\hat{T}\psi(3\Delta x) \\
\hat{T}\psi(4\Delta x) \\
\vdots
\end{pmatrix} = \frac{\hbar^2}{2m\Delta x} \begin{pmatrix}
2 & -1 & 0 & 0 & \cdots \\
-1 & 2 & -1 & 0 & \cdots \\
0 & -1 & 2 & -1 & \cdots \\
0 & 0 & -1 & 2 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix} \begin{pmatrix}
\psi(\Delta x) \\
\psi(2\Delta x) \\
\psi(3\Delta x) \\
\psi(4\Delta x) \\
\vdots
\end{pmatrix}
\]

Students were then asked to calculate the eigenvalues and eigenvectors of the $\hat{T}$ matrix and to observe what happens when the size of $\Delta x$ is decreased. Because the ISW has a defined length, decreasing the step size $\Delta x$ will increase the number of steps required to iterate over the entire box, resulting in a $\hat{T}$ matrix and a wave function column vector with more entries.

### 2. Video Clip Description

The video clip Mary watched during the interview opened with the observer asking Mary and Dave how the length of the ISW, the size of $\Delta x$, and the number of entries in the wave function column vector correspond to each other. The pair then looked at their wave function column vector composed of $\psi(\Delta x), \psi(2\Delta x), \psi(3\Delta x)$, and so on. They noticed that the number of entries in that vector did not change as they varied the step size $\Delta x$ or the length of the square well. Mary was confused as to why that should be the case and Dave attempted to explain why through verbal descriptions and drawings.

### B. Analysis

We used an instrumental case study to analyze Mary’s interview responses. In instrumental case studies, a particular case is used to provide insight into a larger issue [7]. The case we present is that of a student, Mary, who has done a computational activity focused on discretization as both a student and a TA. We present two conceptual challenges that students face when introduced to discretization identified by Mary during the video elicitation interview. This case will give further insight into the challenges students face in transitioning from working with discrete quantum systems like spin-$1/2$ particles to working with continuous quantum systems like the ISW.

Mary may be considered an extreme case [8] or an unusual case. She was part of a larger research project on the computational lab course and happened to later become a TA for that same course. She saw the course material multiple times with new coding and physics expertise being developed between taking the course as a student and being a TA for the course. This gave her more experience with the kinetic energy operator activity and a perspective that most students in the course would not have. Consequently, Mary was particularly valuable to interview because she gave the authors insights into the challenges of discretization that most students would not be able to articulate.

### III. RESULTS: CHALLENGES OF DISCRETIZATION

The purpose of this case study is to highlight some specific struggles students face when engaging with discretization in a quantum mechanics computational activity. During the interview, Mary identified two challenges of discretization. The first was centered around the difficulty of recognizing that functions can be represented as a column vector with a finite number of elements. The second was centered around students’ interpretation of $\Delta x$. She identified both these challenges to be productive for her understanding of wave functions and discretization.

In the following presentation of the challenges, all of the quotes are from Mary during the Winter 2022 (W22) interview rather than from the Winter 2021 (W21) observational video clip of her as a student.

#### A. Functions as Discrete Column Vectors

Of the four computational quantum mechanics activities, the kinetic energy operator activity was the first that required students to recognize wave functions can be represented as a column vector. Mary identified this as a challenge of discretization. During the interview, Mary stated that “...it’s hard to think about what it even means to have the wave function in a matrix form and like have an operator in matrix form.” Later in the interview, she described how the code she and Dave wrote approximates a wave function.

Mary: “Yeah, so we’re approximating a continuous psi, which is an eigenfunction. We’re approximating it with a discrete representation and rather than having some infinitely long set from like — er, yeah, infinite breaking up from 0 to L, because it is a particle in a box, it’s not from negative infinity to infinity. But like we’re breaking it
up into — we’re only taking certain points along the line.”

She described that typically the wave function has a defined value at each position along the length of the ISW, resulting in an uncountable infinite collection of coefficients. In the discretized case, however, she stated that the wave function is being approximated by sub-sampling the infinite set of coefficients. This allows the wave function to be represented by a finite set of coefficients. In the W21 observation video, she reached the same conclusion, but she more clearly articulated the approximation aspect of this during the W22 interview.

Part of the activity required plotting the eigenfunctions of the kinetic energy operator. Mary had identified this task as “intuitive” in the following description.

**Mary:** If you fed in np.arange from 0 to L and just any number of x, you wouldn’t have enough values for your eigenfunction to plot versus, so it would run an error on the code. That is very unintuitive for someone who’s like ‘Well, I should just be able to plot a function, and if you just give it any x value it should be able to spit something out.’ There’s just a lot of elements to it that are confusing I think.

Because the wave function and kinetic energy operator had been discretized in the code, the calculated eigenfunctions had to be plotted against “certain points along the line”. She described this as being unintuitive because typically wave functions have a value at every point along the length of the ISW, but the discretized wave function does not. In both the W21 observational video and the W22 interview, she found the challenge of understanding a wave function as a column vector productive to her understanding of discretizing a wave function. In particular, she noted her experience as a TA for the course helped her recognize that the size of \( \Delta x \) corresponded to how well the wave function was being approximated. She did not remember learning this in W21.

**B. Interpreting \( \Delta x \)**

In the W21 observational video, Mary was confused about why the number of elements in the discretized wave function changed with the value of \( \Delta x \). Her partner, Dave, then drew a representation of the system on their shared virtual whiteboard, shown in Fig. 1. During the W22 interview, Mary was asked about her thoughts on Dave’s drawing. She drew a plot on the whiteboard as shown in Fig. 2 and said:

**Mary:** “I think just seeing \( \Delta x \) on a page is nice, because then you’re like, ‘Okay, what was \( \Delta x \) supposed to be?’ Um, except for that what I’m learning now is that \( \Delta x \) isn’t actually like on a bar graph where you have between two points [sic]. It’s literally just a point on the graph.”

In this quote, Mary reflected that Dave’s drawing was helpful because it labeled what \( \Delta x \) was in relation to the length of the ISW. She then identified two interpretations of \( \Delta x \) that could impact one’s understanding of the discrete wave function. The first was an understanding of \( \Delta x \) as the space “between two points”, as reflected in Fig. 1. She compared this to a bar graph, where the bars occupy a space between two points. The second interpretation she offered, as drawn in Fig. 2, was that \( \Delta x, 2\Delta x, \) etc. were single points with uniform spacing along the length of the ISW. In the W22 interview, Mary shared that as a TA she saw some students in the computational lab course use the “space between points” (SBP) interpretation and others use the “single point” interpretation. Mary noted that she had the SBP interpretation of \( \Delta x \) in the W21 observational video, but after an additional year of experience she developed the “single point” interpretation.

The SBP interpretation of \( \Delta x \) had implications on how Mary and Dave thought about other parts of the activity. Specifically, it affected how they understood \( \psi(\Delta x) \) and how they graphically represented the ISW. Even with these implications, Mary reflected that the SBP interpretation was pro-
ductive for her in the W21 observational video because it helped her visualize the problem and connect the size of $\Delta x$ to the number of elements in the discretized wave function.

The SBP interpretation of $\Delta x$ introduced challenges to understanding $\psi(\Delta x)$ as a value of a function. During the interview, Mary said the following:

Mary: “[The instructor] just showed this in class where you have $\psi(\Delta x)$, $\psi(2\Delta x)$, $\psi(3\Delta x)$ and so on. And that would map to a discrete point on the curve [underlines $\psi(\Delta x)$ in purple and marks $\Delta x$ on the plot in Fig. 2]. Like this is $\Delta x$ and then this’d be some $2\Delta x$ [underlines $\psi(2\Delta x)$ in orange and then marks $2\Delta x$ on the plot in Fig. 2]. So then, I guess, I didn’t understand...why a discretized representation even worked. Like it isn’t a one-to-one, it’s just an approximation of the curve. I don’t think I had any understanding of that.”

She reflected that in W21 as a student with the SBP interpretation, she did not fully understand how to coordinate the discretized wave function with a plot of the wave function vs. position. With the understanding that $\Delta x$ is a space between points, it was not clear to her in W21 how to evaluate $\psi(\Delta x)$. After experiencing the activity for a second time, in the W22 interview, Mary noted that the “single point” interpretation of $\Delta x$ helped her understand that $\psi(\Delta x)$ is the wave function evaluated at a point.

Mary did, however, reflect that the SBP interpretation was productive for her and her partner. As shown in Fig. 1, Dave’s drawing of the ISW labeled $\Delta x$ as the space between points. In the W21 observational video, the drawing helped Mary understand “breaking up L into $\Delta x$” and how changing $\Delta x$ would change the number of entries in the discrete wave function. Mary was asked to reflect on the drawing during the W22 interview. She explained that even though the drawing was not completely physically accurate, it did help her in W21 to understand the connection between $\Delta x$ and the discrete wave function. Additionally, she noted the characteristics of the drawing that might be inaccurate or confusing for students, such as the x- and y-axes and the box not being aligned with the origin.

IV. DISCUSSION AND CONCLUSION

We have presented two challenges of discretizing wave functions in a quantum mechanics computational activity as highlighted by a participant who has experienced the computational activity as a student and as a TA. The kinetic energy operator computational activity was productive and valuable in light of these challenges, as it encouraged students to engage with some of the nuanced ideas of discretization.

The first challenge we identified was in representing functions as discrete column vectors. A wave function, an uncountably infinite collection of points, can be approximated with a finite set of points. Recognizing that wave functions are, by definition, a collection of points may be one of the reasons that the idea of discretization was challenging. According to an analysis of the structural features of quantum notations [9], wave function notation exhibits a low degree of individuation while matrix notation exhibits a high degree of individuation. Individuation refers to how well the important features of an object are represented as elemental pieces of a whole in a given notation. The uncountably infinite set of points comprising the wave function is obfuscated, which makes it more difficult to see how a function could possibly be represented by a column vector.

The second challenge was interpreting $\Delta x$. Some students interpret it to be the space or gap between points rather than a point. This led to some difficulty in understanding $\psi(\Delta x)$ as a value function and manifested in the graphical representation of the ISW. Mary mentioned that the course instructor drew a plot similar to the one in Fig. 2 that highlighted the connection between the $\Delta x$ in the column vector and in the plot of a wave function. She said that this was helpful for orienting students to the activity, but it did not completely eliminate some students from understanding $\Delta x$ as the space between two points. The “single point” interpretation of $\Delta x$ was particularly productive for Mary as a TA because it allowed her to coordinate the discretized wave function with a graphical representation of the ISW.

We believe this study has further implications for highlighting the ways that continuous quantum systems like the ISW relate to discrete quantum systems such as spin-1/2 particles. Activities centered around turning a continuous object into a discrete object may illuminate the ways that common calculations like inner products, probabilities, and time evolution are isomorphic for discrete and continuous quantum systems. For example, writing Python code that calculates the inner product of two wave functions would require writing an integral as a sum, resembling an inner product for a spin-1/2 system.

We offer some implications for instructors, as informed by our study. Terms like $\Delta x$ may carry connotations from previous courses which can affect students’ interpretations of them in a new context like a computational course. Additionally, providing multiple representations of wave functions (graphs, code, function representations, matrix representations, etc.) encourages students to consider the role of $\Delta x$ in discretization. The study also provides evidence that students may benefit from additional instruction on the similarities of functions and vectors to facilitate students’ understanding of how a wave function can be represented as a column vector.

ACKNOWLEDGMENTS

This work is funded by the NSF DUE 1836604 and 1836603.


Using a volunteerism framework to understand the motivations of university students who facilitate informal physics programs

Bryan Stanley
Dept. of Physics and Astronomy, Michigan State University, 567 Wilson Road, East Lansing, Michigan, 48824

Kathleen Hinko
Dept. of Physics and Astronomy, Michigan State University, 567 Wilson Road, East Lansing, Michigan, 48824 and
Lyman Briggs College, Michigan State University, East Lansing, Michigan 48824

Many university students who participate in public engagement do so voluntarily through informal physics programs. The motivations and reasoning students have for volunteering in the first place and continuing to volunteer vary. For some students, volunteering in informal physics programs can influence their career path after leaving the program. For this research study, we interviewed university alumni about their past experiences as student volunteers and the career path they have had since leaving the program. We apply a volunteerism framework to our data to 1) see how it maps onto an informal physics context and 2) understand the evolution of volunteers’ motivations from their initial informal physics volunteer experiences to their present lives. As this study is meant to test the volunteerism framework, we will focus on alumni who continued into education-related careers given that informal physics programs are educational in nature.
I. INTRODUCTION

Physicists and physics students engage with public audiences through a variety of informal physics programs and activities like school visits, summer camps, open houses, public talks, after school programs, and more. Participation in these activities are often voluntary for both the participants and the facilitators. Many volunteers in informal physics programming are university undergraduate and graduate students [1, 2]. Some students volunteer repeatedly and for many hours in these spaces throughout their student careers, even though they are often not earning course credit or being paid [3–6]. We are working to understand the nature of the deep and meaningful experiences that students are having. Studies have looked at the impacts on university student volunteers and have found how some structures of informal physics programs support physics identity development, sense of belonging, and career-related skills [6–12]. Students have different career paths, some of which are education related, and their experiences from volunteering can impact the careers they pursue.

Some studies have looked at the motivations of university students who volunteer in informal programs, noting themes such as positive experiences, sharing science, professional development, and teaching opportunities [12, 13]. In particular, some studies have used community of practice [8–11, 14] and personas [15] frameworks to better understand the involvement of university students volunteers in informal physics programs. However, many of these studies analyze students as they are actively involved in their volunteering and do not look at the careers and long term impacts. Rothman et al. [6] found that current and former student volunteers reported an increase in motivational beliefs. Those motivations have strong ties to student excitement and development of interest and skills that are relevant to becoming a physicist [6].

Here, we interview alumni at various points in their professional careers about their past experiences volunteering in informal physics programs and how those experiences impacted them and their career paths. Our overarching research focus is on the different ways that informal physics programs can affect careers of volunteers. Our initial step in this preliminary study is to apply a volunteerism motivational framework to 1) see how it maps onto the informal physics space, and 2) understand how volunteers’ motivations evolve from their first volunteer experiences in informal physics to their present lives. Before applying to our larger data set, we first apply this framework to alumni whose careers ended up in educational fields. For the scope of this paper, we aim to answer the following: How do informal physics program experiences affect volunteers who go into careers related to education?

II. VOLUNTEERISM FRAMEWORK

For many facilitators, informal physics is not a part of their main job or responsibilities. For example, faculty may have teaching or research responsibilities while students may have coursework or other involvements. The work that most of the facilitators do in their informal physics programs is in addition to their main job or position [2]. We use the term student volunteer to describe the university students who are contributing to the functionality of these programs. In some cases, but not all, university students may be compensated for their involvement in their program, however, we still label them as a volunteer as their main position is being a student and their informal physics involvement is on the side.

Given that participation in informal physics programming is often in addition to one’s current job and responsibilities, there must be some motivations for the volunteer to 1) volunteer in the first place, and 2) to stay in that volunteer position. The volunteerism framework aims to understand those two points for volunteers more broadly. Clary and Snyder define six categories that describe motivations for volunteering (Values, Understanding, Enhancement, Career, Social, and Protective), as described in Table 1 [16, 17]. Pulling from foundations in psychology, this framework was tested with volunteers in public health, hospital programs, psychology programs, and business, with populations of volunteers being non-students, students getting course credit for their service, and students not getting course credit [16]. While this framework is aimed for volunteering more generally, Clary and Snyder acknowledge that these categories may appear differently based on the type of activity the volunteer is participating in [16]. This framework has been used in scientific spaces, such as citizen science programs and environmental sciences, but was adapted with some contextual renaming or some context-based additions [18–21].

For this pilot study, we want to test this framework in the context of informal physics programs to determine if it captures the motivations of student volunteers. Applying this framework to interviews with alumni helps to understand how those motivations connect to the volunteers’ career paths. Because we are testing the framework, we are applying it to a subset of our data. Given that informal physics programs are inherently educational, we first apply this framework on alumni who had a career in education.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Career</td>
<td>Preparation and experience for career-related endeavors.</td>
</tr>
<tr>
<td>Enhancement</td>
<td>Increasing positive feelings of oneself.</td>
</tr>
<tr>
<td>Protective</td>
<td>Reduced negative feelings from personal challenges.</td>
</tr>
<tr>
<td>Social</td>
<td>Building and strengthening relationships with friends and new people.</td>
</tr>
<tr>
<td>Understanding</td>
<td>Gaining and implementing new skills and knowledge.</td>
</tr>
<tr>
<td>Values</td>
<td>Belief that the person finds important to their life.</td>
</tr>
</tbody>
</table>
III. METHODS

For data collection, we contacted the lead facilitators of two informal physics programs: Traveling Physics and After School Physics. Both programs are housed at different large research universities. Traveling Physics is a travel program that visits schools with hands-on physics experiments that were built by its team of primarily undergraduate student volunteers. After School Physics is an afterschool program that partners with multiple nearby middle schools and high schools. A team of undergraduate and graduate students and postdocs visit each school once a week to lead hands-on physics activities and explorations.

We contacted these programs about our study because 1) both programs have been around for decades, 2) both programs stay in contact with their alumni, and 3) the authors have previous involvement with these programs. Program leaders shared information of our study to their alumni contact lists in order to recruit participants to be interviewed. Connections with these participants snowballed to members from two additional programs who were also interviewed. In total, 25 interviews with alumni were conducted.

The approximately 30-minute interviews were semistructured. Some interviews were conducted in person and some were done virtually on Zoom. We asked participants about their past experiences in their program, their career after leaving their university, skills important to their current work, and memorable experiences the participants had in the program, with other volunteers, and with audience members. For analysis, both authors independently coded the interviews with the volunteerism framework using the qualitative analysis program MAXQDA. Code units were full sentences in which the response met the definition of a framework category. A code could be multiple consecutive sentences, but not partial sentences. Sentences could include multiple codes. After each interview, the authors compared codes, to which there was a high agreement, discussed any discrepancies and documented common themes found within each category.

For the scope of this paper, we are only analyzing participants who volunteered in either Traveling Physics or After School Physics and who ended up in an educational-based career, whether that was formal, informal, or a combination of both. We will present each participants’ trajectory linearly, highlighting the different motivational categories that manifest throughout. The three participants we will discuss are “Amber,” “Mark,” and “Claire.”

Amber began her post-secondary career as a student at a two-year college. After taking an astronomy course, she decided to pursue a degree in physics at a large four-year research university where she volunteered at Traveling Physics. Her job trajectory went in many different directions. She worked as lab coordinator, worked in the tech industry, returned to school to pursue a teaching degree, left the teaching program to open up her own business, returned to the tech industry, then hired as assistant director for Traveling Physics.

Mark earned his bachelor’s degree in physics. As an undergraduate, he worked as a teaching assistant for the undergraduate labs. He went to graduate school to continue studying physics. As a graduate student, he volunteered at Traveling Physics. After graduate school, he worked at a university as a lecture demonstration specialist, where he teaches some classes and still participates in public engagement activities.

Claire was a neuroscience major as an undergraduate who wanted to go into education since early college. As an undergraduate, she volunteered in After School Physics. She is currently a middle school science teacher.

IV. APPLYING FRAMEWORK TO THREE CASES

Here, we will discuss the job trajectories of Amber, Mark, and Claire from when they volunteered in their respective informal physics programs to their current position. We present each persons’ story individually and chronologically, highlighting key instances where the participant mentioned one of the six motivational categories. We find that all six motivational categories are present across these three interviews.

Amber: As an undergraduate, Amber got involved at Traveling Physics because her tutor was a volunteer there and he told her that he thought she would enjoy volunteering with the program. Prior to volunteering in Traveling Physics, Amber had experience working with kids through nannying and daycare. Building from her enjoyment of working with kids, one of Amber’s original motivations for volunteering at Traveling Physics was because, “I get to come in, inspire some folks, and maybe get some young people interested in doing science.” Values like wanting to inspire young people is one of the motivators for Amber to get involved with volunteering, but additional categories also help retain her in the program.

As a student, Amber had jobs in retail and pizza delivery; however, “I was doing doing jobs that were just doing jobs.” She had Protective motivation for transitioning to Traveling Physics to reduce the negative feelings she had working at those other jobs. Instead, she wanted a job where she could use “my skill set that I thought I was good at. And so by joining [Traveling Physics], I could just concentrate on school but also enhance my physics learning. So I could go into, let’s say, my E&M class, and I’m struggling with this electricity and magnetism concept, and I could come back into [Traveling Physics] and say, ‘How does this work?’... So now I can move more into a career trajectory, versus I’m just doing jobs to pay the bills.” Traveling Physics was providing Amber with Understanding motivations by helping her to build upon her physics knowledge and skills. In addition, she sees these skills and knowledge to be more helpful in lining up with her Career trajectory compared to her other job experiences.

After earning her bachelor’s degree, she ran the undergraduate labs before getting a job in the tech industry. After being laid off, she returned to her alma mater to pursue a teaching degree. During this time, she worked again at Traveling Physics. Based on her past experiences in the tech industry, Amber came into Traveling Physics with some additional
motivations, specifically with her Values of inspiring women. “For me personally, because I am a woman in science and something I didn’t come to until I was in the tech industry, and I saw how few women there were that I worked with, my personal goal was to go out and inspire young women. So showing them and enabling them that they can do science.” After a semester of education courses, she decided that she did not want a formal educational career; however, she had built Social relationships with the other volunteers. One of these relationships led to her and another volunteer opening up a business together. Amber eventually left that business these relationships led to her and another volunteer opening up a business together. Amber eventually left that business.

Amber says that when she was back in the tech industry, “I was continually really trying to get back into [Traveling Physics] because I realized that was the thing I loved in life. I mean, even my coworkers would say, ‘You go do this thing’ I was continually really trying to get back into [Traveling Physics], and all you do is talk about it. Why don’t you go get a job [there]?’” Her love for Traveling Physics would be an Enhancement motivation, but it was one that developed over time and played a role in her wanting to return and stay involved. Not only was her love of Traveling Physics apparent to her, but her tech colleagues noticed and commented on her enjoyment of the program, even to the point of encouraging her to pursue the program professionally. She applied for multiple positions at Traveling Physics and ultimately was hired as an assistant director.

In our interview, Amber reflected on how her past volunteering experiences impacted her after graduating. While she learned some physics knowledge and skills by building experiments, she also says that volunteering “enabled me to be okay to continue to fail. It was okay to come in here and work on a project for an entire semester, and the project doesn’t work, and that’s okay...Gosh, when [my business] imploded, and we weren’t making any money and that was it, I was like, ‘I don’t know what to do,’ And I’m like, ‘Gosh, what would [Traveling Physics director] do?’ It’s like, ‘Oh, you learn.’ You pick up the pieces. You see what’s the thing you enjoy, and you use that as your guiding light to go forward.” Volunteering helped Amber to gain life skills in overcoming failure, and she used that Understanding later on in her career.

Mark: As an undergraduate, Mark was asked by the physics department to teach some of the introductory physics labs as a teaching assistant. This was one of the early moments when Mark learned that he enjoyed teaching. When he attended graduate school, the Enhancement from his undergraduate teaching motivated him to get involved in the introductory labs and working with the Traveling Physics coordinator. “[In undergrad], I knew how the lab ran, so I helped tell the teaching assistants the next time what we were doing and I would help set up the labs for everybody, started editing the manuals and realized, hey, I like this. So that’s why when I saw that when I got to [graduate school] with [Traveling Physics director], I was very willing to jump in and help and stay in that aspect of it.” In the introductory physics labs, his responsibilities included managing the lab spaces, teaching some of the labs, and leading the teaching assistants.

Mark’s motivations for volunteering were more than just feeling good. “I love physics. I love the doing of it, but if I couldn’t share it with somebody in some fashion, I don’t think I’d still be in it. That’s what I realized I liked, and I felt like I could do that if I was with teaching assistants, the lab students, and, and then ultimately outreach...I feel like I’m sharing it, not dispensing information.” A couple motivations overlap here. He Values the sharing of physics versus the dispensing of knowledge. He also determined that in order for him to stay in a physics Career, sharing physics with others was a necessary aspect for him. Volunteering helped him to Understand how to learn and share physics. “It changed the way I approach physics when I’m learning it. Because...when I’m learning it, I’m trying to make connections now...and help [others] see how it works. The concepts and the principles rather than just here’s some problem solving skills to get you through and graduate. Because that was more the drive of my undergraduate career. [Traveling Physics] was the flip. It was the reverse.”

Mark also had Protective motivations from not feeling trusted to do work in research labs. “In two research labs...they hardly gave me any responsibility or significant responsibility or asked me to do any. I didn’t feel like I was helpful, part of a team or useful. Joining [Traveling Physics], all of that switched. They were like, ‘hey, would you like to work on this or what would you like to work on? OK, pursue it.’ And then they gave guidance, but you got to explore with it and learn and try and maybe you came back and it was wrong and they give you advice and you go work on it some more.” Even when he was taking other coursework and labs as a student, Mark found the manuals and activities to be “cookbook” and that his role was more about “going through the motions. You were just another cog in a wheel...I’d never felt that in [Traveling Physics]. [I] always felt useful, appreciated, helpful.” Traveling Physics provided Mark a place where he felt trusted, able to contribute, and able to learn and apply his skills.

As a lecture demonstration specialist, he currently does demonstrations for physics courses and sometimes teaches courses. He also facilitates teaching workshops with K-12 teachers and does public engagement with K-12 schools. Mark reflects on how Traveling Physics impacted the Career path that he took. “I just want to emphasize again it did steer me [and] had a huge influence on where I’m at now. The outreach, the teaching, the trying to relate more to the public and the students rather than just going into research...I’m about teaching first and foremost. Everything I do is about that now.” As a form of Enhancement, Mark says “[Traveling Physics] made me who I am and I’m grateful for it.”

Claire: She was a neuroscience major who also studied sociology and science education. Since early college, she had been interested in pursuing an education career. This motivation stemmed from her Values of impacting kids and getting them interested in science. “I wanted to have more of a significant like role in in kids lives. And I also wanted to
incorporate my like love of science and so science education really kind of just put all of those together.” She saw an advertisement for After School Physics and joined because of her Career goals. “I’ve had that interest in education for a while. Especially trying to combine education, like with my own passion of science, I think I saw a poster for it and then I saw some people I already knew, like TA’s and LA’s also attending, and I wanted to check it out and I made a lot of cool friends with the group and some cool connections as well.” Here, Social was another motivation for Claire. She already knew people in that particular program and then she continued to make friends and connections through volunteering.

Claire describes her overall experience volunteering in After School Physics as “overwhelmingly positive” and that “it was really heartwarming to play with the kids.” These positive experiences are a form of Enhancement. Some of these positive feelings came from her Social relationships with the other volunteers, noting that “It was really cool to connect with other people that were interested in science education.” With a combination of student volunteers of varying academic ranks, many of whom were physics majors, Claire mentioned that “there was another girl that attended with me. Another undergrad. She was in like astrophysics, I think, but we really connected over being the younger ladies in [After School Physics]. I think we became really good friends.”

Even though Claire was a neuroscience major, she says that in her carpool rides to schools with the other volunteers that they would have a mix of science-related conversations. “It was really cool to just have, like, science discussions with people that can, like, keep up even if they’re not necessarily in the same discipline. But I felt like it was a lot of support.”

As a current middle school science teacher, Claire notes that education was already the Career path that she was going down. “[After School Physics] was one of the first moves I made to solidify myself in sort of a science education pathway. And so as I was exposed to pedagogy there, I was just starting to be exposed to it within my own courses. But it solidified the age group I wanted to work with. The subject matter I wanted to work with. And I think in some ways the demographic I wanted to work with as well.”

### V. DISCUSSION

Amber, Mark, and Claire represent three different types of educational career pathways. Within the three interviews, we find all six categories to be present; however, for each person, some motivational categories were more prominent than others and manifest in different ways. For example, Amber and Claire had some form of Career motivation but in different forms. Amber’s Career motivations began more with physics knowledge and skills which then evolved into pursuing education, while Claire’s Career motivations in education were solidified and narrowed down. Mark had prior teaching experiences that he found positive, but it was those in combination with his Protective motivations from his physics research lab experience that contributed to his pursuit of a career in education. Table II serves as a brief summary of how each category was present across the three interviews and which categories were most prominent for each person. The boldface highlights how that motivation connects to their current work.

There are several limitations with this study. In addition to the participants self-reporting their experiences, most of them are years removed from their time volunteering as students, so recollection of memories should be taken into consideration. In addition, while the alumni we discussed here ended up in education-related jobs, that is not the case for all alumni. We are encouraged by how well the framework illustrated the nuances in people’s motivations, and how those motivations evolved throughout their time in the program and into their careers. Similar to other studies that adapted the volunteerism framework to their own contexts [18–21], we find that this framework appears to appropriate for the informal physics context. Next steps include applying this framework to our larger dataset, which includes a variety of job trajectories, some of which include non-education focused pathways.


Defiance in the face of adversity: a qualitative study of women’s attrition from and persistence in physics

R. Smith Strain
Department of Civil Engineering, Auburn University, Auburn, AL 36849

Matthew Shepherd
Department of Physics, Auburn University, Auburn, AL 36849

Eric Burkholder
Department of Physics & Department of Chemical Engineering, Auburn University, Auburn, AL 36849

Women continue to be underrepresented in physics despite recent developments focusing on improving representation and equity within the discipline. Several recent papers have investigated the experiences of both white women and women of color within the discipline to understand the issues that most often affect them, and how those issues intersect. What has been less clear from previous research is how these factors contribute to women’s decision to stay in physics or leave for another discipline. Much of the literature on persistence has focused on leaving STEM majors for non-STEM majors, but this does not include the possibility of a woman leaving physics for another STEM discipline. In this study, we used a social constructivist lens to investigate the experiences of 5 women in physics at a large research university. We focused specifically on how the women’s experiences contributed to their decision to stay in physics (some even chose to pursue a Ph.D.) or leave physics for another STEM discipline. We find that lack of support, poor treatment by male peers and professors, and interest in physics are all major factors in these women’s decision-making process, but that lack of support was most prominent in the decision to leave physics. This provides interesting counter-narratives of women defying adversity in pursuit of their deep interests in physics but also highlights the need for structural changes and action by male physicists to better support women.
I. INTRODUCTION

In physics, women remain a minority of the student population, only making up about 20% of the population who graduate with a bachelor’s degree in physics [1–3]. Representation of women decreases as they move to higher levels within the academic system. Prior studies have shown the existence of sexism in the current physics culture [4], which contributes to the continued underrepresentation of women in the field. Other studies have shown that women are often perceived as less competent than men in many science fields [5–7], which may lead to overt mistreatment by peers and contribute to women leaving science.

The phenomena of student retention and attrition have been studied most thoroughly in the last 50 years [8], but there is still much room for the experiences of women in science, technology, engineering, and mathematics (STEM) fields to be studied. Talking about Leaving revolutionized the conversation surrounding the issue of women and students of color leaving STEM majors in 1997. This study was revisited in 2019 with quite similar findings, highlighting why underrepresentation and attrition continue to be issues nearly 30 years later. The authors found that although the rate of switching out of STEM majors had decreased since the publishing of the original 1997 study, the reasons for switching generally remained the same (poor learning environment, coursework overload, trouble finding auxiliary resources). They also found some new reasons contributing to attrition, such as a lack of confidence, discomfort with a competitive environment, and having more career options [9]. Talking About Leaving is an ethnographic qualitative study (though it does have a large number of participants, ∼ O(100)). Quantitative studies, such as that by Sonnert et al. [10] have added more understanding of which of the factors identified in Talking About Leaving might be most influential in students’ decision-making. Utilizing data from 499 institutions in the biology, engineering, and physical sciences, they found a disparity between women’s and men’s GPAs (grade point averages) that could possibly be contributing to higher attrition rates for women. This study highlighted the relevance of the institutional environment for the relative performance of male and female undergraduates and also suggested the possibility of multiple factors working together such as institutional location and Carnegie classification.

Though much research has been conducted pertaining to women’s experiences and persistence in STEM more broadly, there is less research on women’s experiences in physics. Recent studies by Santana and Singh suggest that students from underrepresented groups often experience a lack of sense of belonging within physics [11]. In a previous study, the authors suggested that due to the hostile environment created by male peers in physics, it is critical for women to create a community amongst themselves and have resources readily available to support this [12]. Though both studies underscore how underrepresented populations feel unwelcome in physics, they did not investigate how these experiences affected students’ decisions to stay in or leave physics. Therefore, in this study, we seek to bridge the gap by examining the relationship between women’s experiences in physics and understanding their decisions to leave or stay in the discipline. We note that prior studies on retention in STEM have mainly focused on why women leave STEM fields in general and have not investigated the factors that may cause women to leave one STEM field (e.g., physics) for another [9]. As a result, our investigations focused specifically on women who had started college as physics majors and explored what factors those women believed contributed to their decisions to remain in physics or leave for another STEM discipline.

The interpretive framework we chose for analyzing our data was social constructivism [13], which posits that multiple realities can exist. This is important when considering the different reasons why a woman may want to stay in physics, and also why she may leave: by understanding that each experience and perception is unique, we conclude that no singular reason can be attributed to why a woman would choose to stay or leave the field. The same logical steps can be taken to understand our use of social constructivism in framing our epistemological lens that knowledge and thus evidence is based on the subjectivity of the individual. By assuming reality is known through individual experiences, we are able to take the women’s reports of their experiences as truth. This framework is more broadly related to feminist standpoint theories, which hypothesize that knowledge is produced by one’s personal experiences and that researchers should highlight the experiences of the oppressed group [14], which includes women in physics. Standpoint theory calls upon the researchers to use the women’s experiences as a tool to dismantle the system of the oppressor. Thus, by taking the women’s experiences and statements as objective truths, we seek to undermine the patriarchal structure of physics.

II. METHODOLOGY

We conducted five interviews with women at a southeastern land-grant university. Four women were enrolled in the physics program at the time of the study, and one woman had left the program (pseudonym Claire). Two of the women who were currently enrolled were PhD students (Beatrice and Alyssa), and two were undergraduate students (Mary and Grace). The participants were recruited through a series of email advertisements sent via undergraduate advising. Each interview lasted 1 hour at the maximum, with each participant receiving a $25 Amazon gift card as compensation for her time. The participants were not required to use their real names or provide any contact information to sign-up for an interview. Additionally, they were given the option of speaking to a white woman or a man of color, depending on who they might be most comfortable with. Both interviewers were undergraduate students to ensure that there was no potential power imbalance between the interviewer and the participant. This was done in the interests of psychological safety [15].
as prior research shows that the environment might influence participants’ responses. The protocol was adapted from Talking About Leaving Revisited [9] (which is a semi-structured protocol) and modified to fit the nature of the interviews. This work was determined to be exempt from IRB review due to the absence of identifying information in the study recruitment and interview process.

All interviews were transcribed by the research team. For each interview, the first and second authors independently generated a narrative summary to understand different points of the woman’s educational journey within her field. These two narratives were then compared and discussed by the first and second authors to identify which themes arose in both narratives and which themes may have been noticed by only one of the interviewers. The first author then used an inductive coding scheme to generate themes that were present across all of the interviews and how each woman placed importance on these themes in her decision to leave or stay in physics. Four of the interview participants were White women and one was a woman of color. For this study, we will focus only on the experiences associated with gender and not on the intersectional nature of gender and race in physics. We made this decision because we only have the perspective of a woman of color who chose to stay in physics and not one who left the discipline. Thus we would not be able to draw the contrasts we would like in identifying factors that contribute to the persistence or attrition of women of color specifically.

The first author is a White woman majoring in engineering at the same university as the participants, and the second author is a Black man majoring in physics; both are undergraduate students. The third author is a White man and faculty in the Department of Physics at this university. All three authors are committed to improving the experiences of those underrepresented in physics and engineering at the university. The first and second authors are able to draw on their experiences as minoritized students in the physics classroom to better contextualize the experiences of the interview participants. The third author provided guidance on the qualitative analysis and protocol development and believes that his position of privilege requires him not only to educate himself about these women’s experiences but to then take action to improve the environment for women in his department.

III. RESULTS

We found four major themes that were present in all women’s interviews: (1) lack of support structures within physics, (2) negative treatment by male peers and professors, (3) acts of defiance against barriers to participation, and (4) suggested solutions to the problem of underrepresentation in physics. We did not originally intend to combine the experiences of graduate and undergraduate students but ended up finding that the themes were quite consistent across the two groups. We believe that the graduate students’ accounts of their experiences serve as an example of how the stigmatization of women extends from undergraduate school into graduate school. By allowing women to tell their stories of their experiences in physics and asking them why they stayed or left, we are drawing on Standpoint Theory and centering the experiences of the oppressed. In addition, we gave the women the opportunity to draw on their experiences and suggest challenges to the prevailing physics structure that could improve the experiences of future women.

A. Lack of Support

Almost all of the women interviewed mentioned a lack of support in the physics major. Claire described a time when she reached out to her advisor for a phone number of someone who could answer questions pertaining to the physics major, especially about being a female in STEM, and her academic advisor (who is a staff member, not a faculty member) said that she did not have one. She went on to say, “That was a really big part in my decision [to switch], is if I felt like, there was no support system set up, so I couldn’t even, like, try to find help for myself.” Claire also mentioned that she felt that she would “not have been given the opportunity to thrive” in her intermediate mechanics course, as the class was graded on a curve, and she felt as though there were not enough resources available to help her do well. She went on to say “I think the expectations can be different at times...well kind of on both sides because then one if you’re a female in these male-dominated classes, the expectation is almost higher on you to like, do better and work harder, but at the same time, I felt like there was a lack of a support system.” Claire mentioned that had there been some sort of support for women in the major, she may have stayed: “When I was becoming unsure about my major, if I had been able to talk to someone, I may still be there today.” It should also be mentioned that Claire’s switch to biology was due to her interest in medicine, and she found that biology both aligned with that interest and provided more support.

The theme of support also appeared in conversation with the women who stayed, with Mary noting that although she did not experience any barriers when seeking support, the number of sources of official support was small. She said her supplemental instruction classes “were fine, but they only lasted so long.” On the other hand, Beatrice, a current doctoral student who had previously graduated from a small liberal arts college in the Midwest, found that her graduate teachers were not an active source of support for her. This caused her to seek more outside support from women who were in the department but outside her immediate cohort.

B. Male Treatment

It is not a novel finding from these interviews that women experience negative treatment from their male peers and professors. However, our research aimed to understand why
women choose to stay in the face of negative male treatment, and when this treatment might compound with other factors to lead her to leave the discipline.

**Treatment by faculty:** Mary described what she felt was an “internal stigma,” describing how “sometimes, [she] felt like...[she] already walked in as if [she] didn’t belong, because [she] was outside the box for what somebody would consider a typical physics undergraduate, just because [she’s] a woman.” She also cites lab classes in which her opinions were ignored. When asked about how she dealt with these situations, she told us she befriended another woman in the cohort above her, “Just having somebody who could relate and make me not feel crazy, because, you know, sometimes I felt like I was overreacting about it.”

Claire said “It’s hard to describe, like, an environment, but, like, going to office hours and asking questions, and just kind of feeling like I was being looked down on as soon as I went through the door...It felt intimidating, at times.” Grace mentioned that her questions towards professors yielded different responses than if they had been asked by a man, and also said that oftentimes her answers on assignments such as homework were correct (while her male peers’ were not), but no one asked her for help. Alyssa (pseudonym) believes that there is a common societal misconception that women take longer to understand and grasp concepts than men. She provided a classic example of when one professor replied with “You don’t know?” when she approached him with a question.

**Treatment by peers:** Mary said “People would make unwanted approaches towards me, trying to like, almost harass me, a little bit,” when talking about the disadvantages of being a woman in physics. Interestingly, on another hand, Beatrice (graduate student), noted that a lot of things that she feels would be perceived as offensive were actually not to her. She referenced a time when a superior referred to her as “sweetheart”, and her peer was shocked that she was not offended. The same woman also went on to say that without physics, she felt she did not have a purpose. Beatrice also acknowledges this in her interview, answering the question about staying with the simple word “Stubbornness.” Beatrice also says that she has always found math to be beautiful and that in other subjects she was the smartest person in the room, while in the physics classroom, she wasn’t and she enjoyed the challenge.

**IV. DISCUSSION**

We found that both women who left physics and stayed in physics commented about the lack of support they felt within physics and negative male treatment to varying degrees. The women who persisted in physics were to be doing so due to

As mentioned previously, a primary aim was to understand why women decided to stay in the face of this discrimination. We have chosen to call these themes “acts of defiance.” The women who had persisted in the face of discrimination all cited their passion for physics as the primary reason they chose to stay in the discipline. For example, Mary knew that she wanted to be involved in astrophysics, and explained that while her freshman year she considered switching out due to the difficulty of the courses, she “could not imagine studying anything else”. Along the same lines, Grace told us not only of the fascination she had with the subject but also that there was a side of her that wanted to prove people who told her to switch majors wrong, telling the interviewers “When people try to tell me what to do, it makes me wanna, you know, [be like] ’Let me prove you wrong.’” The same woman also went on to say that if she had not have a purpose, Beatrice also acknowledges this in her interview, answering the question about staying with the simple word “Stubbornness.” Beatrice also says that she has always found math to be beautiful and that in other subjects she was the smartest person in the room, while in the physics classroom, she wasn’t and she enjoyed the challenge.

**D. Solution**

The final question in the interview protocol pertained to the future of women in the physics department, asking the participants what they would change about the program and what advice they would give to the department about retaining and recruiting more women. A common response was simply more outreach, with many participants saying that the major continues to be smaller due to less outreach than other STEM majors. Claire mentioned that she believed weed-out classes should not be as difficult if the department wants to retain people in physics, and spoke of the difficulty of courses and how that can deter people by making them fear that higher-level classes will be similarly difficult and inaccessible. Claire mentioned that more resources and programs, such as a mentorship program, would be beneficial to the major to not only guide undergraduate students but to facilitate camaraderie among peers. A prior study [16] suggests that required courses covering implicit bias and discrimination in physics may be beneficial for creating a welcoming environment for women.
a deeply-seated passion for physics [17] and to defy expectations, whether the expectations of their peers, professors, or even families. The women we interviewed suggested that providing more formal support within the major, improving outreach, and reducing barriers like weed-out classes could be potential ways to try and improve the representation of women in their department.

The theme of negative treatment by male peers and professors has been a recurring theme not just in the interviews conducted but for years in the literature in STEM fields. These male behaviors include a lack of support, disengagement by male professors, and implicit or explicit sexism by male peers and instructors. Prior literature suggests that these behaviors are common in physics and other STEM disciplines [6]. Dancy & Hodari [18] found that, while men might be aware of these issues in the abstract, they often use discourse that distances them from the issues. They might make claims such as “not in my classroom,” or argue that changing sexist attitudes and behaviors is beyond their control. For example, one professor in [18] stated that he speaks to his class at the beginning of each semester and emphasizes that he understands how this environment is for women, yet he forgoes the efforts of trying to make the environment better.

One of the ways that male faculty and peers often fail to provide support for women in physics is by not recognizing women as “physics people.” [5]. The importance of this recognition has been continually shown to be related to outcomes for women in physics, such as self-efficacy [19], sense of belonging, and persistence. The failure to recognize this often leads to discrediting women and is harmful to diversifying a homogeneous community—seeing that the perception of those in physics careers are usually white men. Recognition as a “physics person” by high school physics teachers seems to be most important for interest and retention in physics, as students and teachers are often able to forge closer relationships in smaller classrooms. By fostering a greater sense of support in the physics department, we have the opportunity to improve women’s sense of belonging, which Hazari et al. have empirically linked with identity development.

The women’s narratives highlighting a deep-seated interest in physics align with Hazari’s findings that interest is connected with career aspirations. Almost all women interviewed were motivated by a passion for physics and frequently mentioned that encouraging interest in physics would be important to recruiting more women physicists. Grace, one of the aforementioned women, explained to interviewers that she enjoyed watching movies that had to do with physics, but “didn’t see anyone in those movies that [made her think] this could be her in here”, which brings in another concern of underrepresentation of women who are in the physics community in pop culture and media. On the other hand, Mary, who said she could not see herself doing anything else, explained that it was her father who piqued her interest in physics, telling interviewers “My dad works on the arsenal [in Huntsville], and he’s really interested in space, so when I was a little kid he taught me a lot about it. I guess it just kind of stuck with me.”

Yet, we saw that the women often mentioned that they were not seen as physics people by their university peers and professors, relying only on recognition of their own academic strengths. Connecting these two lines of prior research suggests that one major step toward increasing the representation and improving the experiences of women in physics could be for men to more proactively recognize women as physics people—particularly physics instructors. We have some preliminary evidence that instructors might be able to foster this recognition simply by adopting more student-centered teaching methods. In addition to this, we also think a mentorship program would be pertinent, as well as having an alumni connection forum for new women in the program to connect with women who have graduated and are following through career paths. This would serve as not only a source of knowledge, but a source of encouragement to push through the major.

V. CONCLUSIONS AND FUTURE DIRECTIONS

In this qualitative study, we interviewed 5 women in a single physics department and found that they commented on a lack of support and negative treatment by male peers. When probed about why they stayed in the department, women reported acting in defiance of negative treatment or low expectations or being driven by a passion for physics that eclipsed these issues. The stories of women who persist in physics provide interesting counternarratives illustrating how women can persist in hostile environments. However, these acts of defiance should not be what is required simply to get a degree in a discipline you find interesting. Physics departments should do more to provide support for women, but as prior studies argue, should also interrogate their own cultures and power structures to see how men in their department perpetuate the marginalization of women through discrimination or simply through willful inaction.

There are some signs of change in the department studied. For example, there are three male faculty (of 20 total male faculty) who have volunteered to serve as mentors and advocates for female graduate students through the Society for Women in Physics chapter. This does not necessarily indicate cultural change, but could simply be the acknowledgment of some individuals that they need to actively work to change the environment for women in the department. We recommend that the department studied provides a structured support network for women in physics at all stages of their careers. In addition, previous efforts have highlighted the need for increased communication and transparency, which may improve the awareness for existing support structures. Efforts for coordinated, data-informed cultural change are currently underway and being studied thoroughly.
Characterizing the ‘design-science gap’ in an engineering design-based laboratory unit in an introductory physics course for future engineers

Ravishankar Chatta Subramaniam (he/him)
Department of Physics & Astronomy, Purdue University, 525 Northwestern Ave., West Lafayette, IN 47907, U.S.A.

Amir Bralin
Department of Physics & Astronomy, Purdue University, 525 Northwestern Ave., West Lafayette, IN 47907, U.S.A.

Jason W. Morphew
School of Engineering Education, Purdue University, 610 Purdue Mall, West Lafayette, IN 47907, U.S.A.

Carina M. Rebello
Department of Physics, Toronto Metropolitan University, 350 Victoria St., Toronto, ON, M5B 2K3, Canada

N. Sanjay Rebello
Department of Physics & Astronomy, Purdue University, 525 Northwestern Ave., West Lafayette, IN 47907, U.S.A.

It is essential to equip the next generation STEM workforce with skills that are crucial to solve real-world problems. The success of this preparation rests largely on the instructional innovations in science classrooms for postsecondary students. Reform documents and research suggest that integrating engineering design (ED) experiences add value to science courses. In this study we explore the ways in which students in an introductory physics course developed a solution to a prescribed real-world problem. By carefully analyzing students’ lab reports, we attempted to gain an understanding into the design science connection in their solution. We coded for five aspects to gain insights into the effectiveness of the scaffolds we had provided to guide the students through the task. Results of this study have implications on how to provide appropriate scaffolds, particularly in ED based tasks to maximize science learning.
I. INTRODUCTION

In preparation towards STEM careers, students need to develop strong understanding of scientific and mathematical principles; think creatively, reflect and iterate on their ideas; and process information from a variety of sources [1, 2, 5, 6]. One way to achieve these goals is by integrating science concepts and engineering design (ED) through real-world problems [5, 7]. Design challenges lend a motivating context for students to learn and apply science concepts. The iterative process of working through a design challenge provides students opportunities to learn new concepts, evaluate their approach and revise their thinking [8]. However, students often apply trial and error and other heuristics to solve a design challenge that does not necessitate the learning of science, resulting in what is called the “design-science gap” [9].

The goal of this study is to operationalize the ‘design science gap’ in an introductory undergraduate physics course for future engineers. Though there may not be a unique way of characterizing the design-science gap, we develop a procedure to gain insights into the depths to which students are able to apply physics concepts to their design. Our study explores the extent to which students use principles of physics in their ED challenge. Our research question (RQ) is:

RQ. With what level of completion and correctness do students apply physics concepts to their ED challenge? What evidence, if any, do we find for a design-science gap?

This study is a preliminary investigation to understand the design-science gap in a calculus-based mechanics course at the undergraduate level.

The results of our study have implications for educators to appropriately scaffold science learning through engineering design.

II. LITERATURE REVIEW

Research has shown that integration of engineering into learning science can provide an authentic context for understanding and application of science [10]. However, students do not necessarily rigorously apply science concepts while solving an ED problem [11, 12]. Despite pedagogical innovations aimed to integrate ED and science learning, students often have difficulties in connecting their design challenge with the underlying science [13]. It is incumbent on educators to bridge this design-science gap. One way this can be achieved is by designing activities with appropriate scaffolds, and pedagogical strategies [14].

In this context it may be worthwhile to consider if science and engineering are indeed different. Radder [15] is of the view that any attempt to define and differentiate science and engineering would be debatable since science and engineering have as much similarities as differences. Meaningful differences between science and engineering are not in terms of their practices (e.g., asking questions, observing, experimenting) but more in terms of their purpose and motives [10]. For the data analysis of this study, we will adhere to the framework provided by the Next Generation Science Standards (NGSS) [6]. Even here it may be noted that NGSS emphasizes the difference between science and engineering in only two of its eight practices.

III. CONTEXT OF STUDY

This study is situated in a large-enrollment, first-semester, calculus-based undergraduate physics course at a large U.S. Midwestern land-grant university. A significant reform in this course over the last two years has been integrating ED into the laboratory component of the course. This course has an annual enrollment of about 2500 students, of which about 85% are engineering majors and the remaining are science majors. Engineering majors are concurrently enrolled in a first-year engineering course focused on ED, which creates a unified engineering design experience for these students spanning multiple courses. Non-engineering majors are provided tutorials on ED prior to starting the ED challenge.

The course adopts the principle-based approach [16] such that the content is divided into three units each focused on a fundamental principle: momentum, energy, and angular momentum. Common threads include a focus on systems thinking, modeling, and making assumptions and approximations. The weekly schedule includes two 50-minute lectures, one 110-minute laboratory, and one 50-minute recitation focused on problem-solving. The laboratory segment had 13 sessions, encompassing three multiweek ED challenges. In weeks 6 through 10, student groups worked on an ED challenge focused on the energy principle. In week 6 they engaged in problem scoping and solution generation. In weeks 7 - 9 they completed inquiry-based activities, which were related to the ED challenge, by using hands-on equipment, VPython code and PhET simulations to help build conceptual and computational models for (i) launching a payload along an inclined ramp using a spring-loaded device (week 7), (ii) dropping a coffee filter or firing a projectile under the influence of air drag (week 8), and (iii) bouncing a falling object off a hard surface with a certain coefficient of restitution (week 9). Each week, groups were asked to revisit the ED challenge and audio record their group discussions on ideas for solving the ED challenge. This study is based on the final lab report the students submitted in week 10.

A. Participants

The participants were 27 student groups of three students each, enrolled in the two laboratory sections for which the
first author was the teaching assistant (TA). The lab groups were formed based on the CATME (www.CATME.org) Team-Maker Survey [17] given to the students at the beginning of the semester. There was no statistically significant difference between the average performance on exams of these groups and the overall class performance.

In this study, we analyzed students’ final (week 10) report in response to a prescribed engineering design (ED) challenge problem. Though students worked in groups, they individually submitted weekly lab reports. One lab report per group was randomly selected and qualitatively analyzed for this study. We investigated the design approaches students adopted, to what detail they were able to justify their design decisions, the science connections they made to the lab activities which were partly intended as scaffolds, how well they were able to justify the assumptions and approximations in their design solutions, and whether they were able identify the limitations in their solution. This analysis would address the overarching question: To what extent did students appropriately use physics concepts and principles in the solutions of their ED challenge?

B. Engineering design challenge

Seven essential ED characteristics [18] guided the development of our challenge. Students worked in teams to first identify the overall context of the problem and then generated possible ideas or solutions using what they knew about the problem as well as using relevant physics knowledge. The teams created and tested their plan, recorded results, and used their current scientific knowledge to explain their design. They shared their ideas and gathered feedback from other teams and the graduate teaching assistant (GTA) and used this information to revise, improve, and retest their original model. Professional development was provided to GTAs to facilitate the design challenge in the lab.

Although the ED challenge was situated in the laboratory component of the course, it was not confined to the lab alone. ‘Expansive framing’ was used to integrate and scaffold it with other learning experiences in the course [19]. In the lecture, the instructor asked students to reflect on how physics concepts presented therein might be relevant to the ED challenge. The weekly recitation problems that students solved tied into the ED challenge. In lab, to facilitate integration of science in ED, we modified the ED cycle [20] by taking a ‘detour’ into inquiry-based lab experiences using hands-on equipment and VPython to investigate the science concepts and reflect how these concepts apply to the ED challenge [21].

The ED challenge problem [21] presented to the students read as follows: “Pristine natural habitats of endangered species such as the gorillas in the Congo River basin are becoming increasingly rare. Today, these habitats and the endangered species that inhabit them need to be not only protected but even sustained by humans. As a member of a team of engineers volunteering for a non-profit organization, you are asked to design a system that can launch a payload of food to an island in the Congo River and land it safely for the gorillas. Each payload is about 50 kg, and it must be delivered to a habitat area located on an island in the Congo River that is about 150 m away from the riverbank. To avoid contributing to global warming, the client wants you to use a means that would minimize the carbon footprint of the delivery. Furthermore, the client also wants to ensure that the habitat remains pristine, so that neither humans nor a robotic machine must disturb the flora and fauna of the habitat while delivering the food”.

IV. METHODS

A. Data collection

Our main data source was the students’ lab reports submitted in week 10, after completion of the ED challenge. We specifically examined students’ responses to four tasks (1 – 4) below.

Task 1. From the previous labs, what are the solutions (or iterations of the solution) you have explored?

Task 2. Based on the solutions/iterations you have, what would your final solution be? Describe in words the ideas you are using from any of the previous labs and how you would combine them into this final solution.

Task 3. Present your final design in a diagram/sketch. Provide a labeled sketch to better explain all possible aspects (physics and engineering design) of your final solution.

Task 4. What are some of the limitations of your final design? Think about possible physics related limitations of your design. Additionally, explore possible limitations that are not related to the physics concepts.

B. Data analysis

Student responses to Tasks 1–4 were qualitatively coded through several rounds of open coding by the first author. Five inter-related themes emerged from our analysis: Physics Concepts (PC), Design Solution (DS), Making Assumptions and Approximations (AA), Recognizing Limitations (RL), and engaging in Iterative Thinking (IT).

We coded for each of these dimensions on four levels of performance. Additionally, five of the 27 reports were independently coded by another trained rater for inter-rater reliability (IRR). Codes were compared and reviewed by the raters until a consensus was reached. The first author recoded the remaining transcripts post the IRR test. The levels of performance for each dimension are described below.

Design Solution (DS): Levels of performance of the design solutions as shown in Table I. We also separately categorized the type of solution to the challenge.
“The final solution involves a payload sitting on a ramp. The payload would be launched using a spring. The properties of the ramp, spring would be used to calculate its initial velocity and the maximum height it would reach. Once the payload reaches maximum height, a parachute will launch to increase its drag force. The payload will fall onto the sandbar located at the shore of the river on the other side, contributing to a lower coefficient of restitution and making the payload bounce less. Landing on the shore would also not affect the wildlife in the area.”

Iterative Thinking (IT): Students were asked to iterate on their design based on the hands-on tasks and simulations that they worked on in the lab, which in turn was related to new concepts and problems addressed in the lecture and recitation. Levels of performance (Table II) were based on the extent to which the design iterations made connections to the lab experiences each week.

Recognizing Limitations (RL): The goal was for students to progress toward the design solution by applying relevant physics concepts. In the process, it was important that students were aware of the limitations of both the design and physics. Levels of performance shown in Table III were used to rate the design limitations.

Assumptions and Approximations (AA): We emphasized to students that real-world problems are complex and encouraged them to simplify the problem by making appropriate assumptions and approximations. Levels of performance shown in Table IV were used to rate the identification of assumptions and approximations.

<table>
<thead>
<tr>
<th>TABLE I. Design Solution (DS) levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Level 0</td>
</tr>
<tr>
<td>Level 1</td>
</tr>
<tr>
<td>Level 2</td>
</tr>
<tr>
<td>Level 3</td>
</tr>
</tbody>
</table>

Here is an example for a Level 3 DS response. The solution approach has details on what materials are used, how and why, along with a justification.

**Physics Concepts (PC):** We expected students to invoke at least the linear momentum and energy principles in their design in some form. Spring force, energy conservation, projectile motion, drag force, collisions and coefficient of restitution are a few of the concepts we expected them to apply. The levels of performance are shown in Table V.

<table>
<thead>
<tr>
<th>TABLE V. Physics Concepts (PC) levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Level 0</td>
</tr>
<tr>
<td>Level 1</td>
</tr>
<tr>
<td>Level 2</td>
</tr>
<tr>
<td>Level 3</td>
</tr>
</tbody>
</table>

Here is an example for a Level 1 response for PC. A set of physics concepts has been merely listed or identified.

“Spring force, momentum principle, energy principle, air drag, collision”.

V. RESULTS & DISCUSSION

To find answers to our research question we conducted a quantitative analysis of the code frequencies which emerged from the qualitative analysis of students’ written work. In particular, to gather evidence for any design-science gap, a McNemar’s Chi-square test was performed to examine the relation between the Design Solution (DS) and the Physics Concepts (PC) used. The test showed a statistically significant difference (p = .013, 2-sided) between the proportions of students who scored high (Level 2 or 3) in DS compared to PS, providing evidence of the design-science gap.

The prevalence of the five emergent themes is shown in Fig. 1.

**FIG. 1.** Performance level on emergent themes in the ED tasks.

The design-science gap is evident in that a larger fraction of groups (19 of 27) scored at the higher performance level (Level 2 or 3) for the Design Solution (DS), compared to a smaller fraction of the groups (11 of 27) for the Physics
Conclusions (PC). A comparison of the codes bears out some interesting trends highlighting the design-science gap, in that a high score for design solution (DS) does not correlate with a high score for physics concepts (PC). This could be due to the nature of the scaffolds provided. In Task 1 we only asked the students to ‘list’ the science concepts related to their design. If we had rather asked students to ‘elaborate and explain’ how the physics concepts would apply to the design, we believe the scores would have shifted to higher levels for PC. This reinforces the importance of structuring our instructional scaffolds well. The data also shows that a high score on the physics concepts (PC), in most cases, leads to a strong design solution (DS), but the converse is not true, which is consistent with literature [6, 15].

A large fraction of the groups (24 of 27) scored at a higher level (2 or 3) for the aspect of Iterative Thinking (IT). This is expected as students were specifically asked to iterate after each lab. Further, a higher score on Iterative Thinking (IT), by the nature of the scaffolds, seems to yield a higher score on the design solution (DS). Although the iterations were regarding their design, it appears that the students were motivated by the new physics concepts they may have learned from the labs.

Only about half the groups (14 of 27) scored at a higher level (2 or 3) assumptions/approximations (AA). This is interesting because assumptions and approximations are highlighted in the Recitation aspect of the course. It appears that students were unable to expansively frame their experiences in the Recitation to the design challenge in the laboratory. The levels on assumptions/approximations (AA) are more closely aligned with level of physics concepts (PC) than the level of the design solution (DS).

A large proportion of groups did well in recognizing limitations (RL) with a significant number (17 of 27) citing both physics and design limitations. Consequently, a high score on recognizing limitations (RL) seems to overlap with a high score on both physics concepts (PC) and the design solution (DS).

Finally, only two of the 27 groups scored a maximum (Level 3) on all aspects, by providing highly detailed ED descriptions and by applying physics concepts. These two groups also drew connections to the VPython and PhET simulations, making it clear that they made the best use of the scaffolds provided.

VI. LIMITATIONS

Though all students submitted their reports, we selected only one report per group. This study does not delve into the variations one may expect within a single group. We believe some students did not quite understand what was exactly expected out of them while responding to the prompts. It is natural that different students interpret terms differently. For example, there is not a unique way in which the terms ‘assumptions’, ‘approximations’ and ‘limitations’ may be interpreted. Presenting the students with a glossary of terms with specific and clear descriptions of the terms involved in the prompts would have helped to elicit more concrete responses. We had asked the students to ‘list’ the various science concepts they may invoke for their design. If we had asked them to ‘explain and elaborate’ instead of merely listing the concepts, we could have obtained a more detailed response from the students. Finally, as much as an understanding of physics concepts is important for a superior design solution, so are mathematics and computing skills. The design solutions presented by our students may have been richer had scaffolds been provided for these aspects.

The results of our study are specific to the given ED challenge. They may vary if a different ED challenge had been provided, or if students were allowed to choose their own ED challenge.

VII. CONCLUSIONS AND IMPLICATIONS

We explored the design-science gap in the context of a multi-week laboratory ED challenge in a first-semester calculus-based physics course. As may be evident, this study is based on only the written reports submitted by the students at the end of the multi-week ED challenge. We are currently working on expanding on this preliminary investigation by analyzing the students’ work in all the weeks of the ED challenge.

We found that while students mostly closely followed the scaffolds provided through the accompanying lab activities each week in inventing a solution to the design challenge, a significant fraction pursued alternative design strategies. This indicates that providing students with scaffolds that may suggest a particular design solution does not necessarily limit their ingenuity in coming up with their own design solution. In addition, we found mixed results on the extent to which students apply their physics concepts to the ED solution. Students who scored highly on the physics concept aspect of their design task also scored highly on their design solution, but not vice-versa, showing evidence of a design-science gap in that the evidence of an effective design solution does not imply a sound knowledge of the underlying physics.

In summary, this study has revealed several interesting leads into students’ understanding of ED and physics. It appears that if students demonstrate an understanding of physics concepts in their design solution, their design is rich and detailed. However, the converse is not supported by the data: strong design solutions do not always reflect a strong grasp of science. It seems that the successful use of ED as a context to learn physics rests on providing students with appropriate scaffolds to integrate their learning of physics concepts into their ED challenge, so as to enhance the design science connection.

ACKNOWLEDGMENTS

This work is supported in part by U.S. National Science Foundation grant DUE-2021389. Opinions expressed are of the authors and not of the Foundation.
Investigating Academic Burnout in Undergraduate Physics Experiences

Harshini Sunil and Bethany R. Wilcox

1Department of Physics, University of Colorado, 390 UCB, Boulder, CO 80309

The term burnout is being used at increasing rates among physics students, particularly in the wake of the changes to instructional modality prompted by the COVID-19 pandemic. Combined with research linking burnout to negative health and career outcomes, this increase in burnout presents a concern with respect to students’ performance and retention. In this research, we will examine the experiences of undergraduate physics students in order to understand how they experience burnout. At the beginning of Fall 2022, we conducted interviews with 7 undergraduates enrolled in upper-division physics or engineering physics classes at two large research universities. We analyzed the data to determine key symptoms that students who self-identified as having feelings of burnout experienced. Following the interview study, we conducted a survey of 24 students to further investigate the symptoms of students that experienced burnout in physics. This paper presents the symptoms of the students who self-identified as burned out and also discusses strategies used by students with lower levels of burnout.
I. INTRODUCTION

With the pandemic affecting students and creating stressful learning environments, burnout is a growing term used by students to describe their experiences in college. While academic burnout is not yet defined as a mental illness, it is a physiological syndrome in response to stressors. The undergraduate physics degree is challenging and stressful, consisting of demanding core classes and electives. The experiences of physics students can involve significant stress and negative emotions, which could lead to feelings of burnout. The purpose of this study is to characterize the experiences of burnout within the physics classroom by analyzing the results of this mixed-methods study with the eventual goal of understanding the ways in which physics students experience burnout. By defining burnout in the context of physics courses, the results of this study can help professors and students recognize warning signs of burnout and lay the foundation to build tools to combat burnout within the physics curriculum. The value of defining burnout in the context of physics undergraduate studies may help mitigate stress-related conditions, improve attendance, and create a positive classroom environment [1].

Formally, burnout describes a state of excessive or prolonged stress that causes emotional, physical, and/or mental exhaustion [2]. When burnout becomes excessive, it can adversely affect both health and academic performance [1]. Students experiencing burnout may feel exhausted due to coursework demands, have a detached or cynical attitude, and/or constantly fear failure [3]. We aim to better understand the challenges and difficulties encountered by students in physics courses at the university level to determine how burnout affects them. Burnout is not limited to students; anyone who feels prolonged stress causing mental exhaustion can experience burnout. Usually, burnout is due to high levels of demand, overworking, lack of self-care, lack of control, etc. [4], elements that are often present in the physics classroom.

II. SIGNIFICANCE & BACKGROUND

Traditionally, burnout has been studied in professional work fields such as healthcare and business [5]. However, students perform tasks that can be considered as work (e.g., going to class, assignments, exams) often with high stakes [6]. Research specifically on academic burnout is limited and has been mostly conducted in other countries. A German study [7] used The Oldenburg Burnout Inventory (OLBI) to study academic burnout and this instrument served as a useful starting point in designing the work reported here. Additionally, some work has been done in China and Korea; however, these studies are mostly written in languages other than English, making them hard to access without translations available. However, the limited existing research expressed concern for academic burnout in highly demanding majors or fields [1]. Burnout can lead to lower commitment, reduced productivity, low morale, lower attendance, and lower human consideration [6]. Physics students often have heavy course loads and many students also have other degrees they are pursuing (i.e., math, astronomy). In particular, 25 of the 31 students in our study had another major or minor, on top of their physics or engineering physics majors. From a study done in Taiwan, students enrolled in multiple programs showed more burnout, which correlated with reduced academic performance [4]. Anecdotally, there is a sense that the pressure for physics students is more acute and rigorous than in many other majors, and this rigor only increases over the course of the 4-year degree.

Similarly, burnout generally results from long-term stimulus and brings a range of physical and emotional symptoms [3]. Emotional exhaustion and cynicism are two core symptoms of burnout, and this can impact motivation, increase feelings of fear, self-doubt, failure, etc. Studies have found that these negative feelings are potentially harmful to students and can cause poor performance [1].

An important distinction to make is the difference between stress and burnout. Burnout is a result of chronic and prolonged stress. Both often overlap in symptoms and feelings, but burnout can contain feelings of hopelessness and a cycle of negative emotions [2]. There is also a distinction from mental illness to burnout. Many mental and physical illnesses share some of the same symptoms as burnout. When analyzing the experiences of students, we must be mindful that other illnesses exist and can impact (or not) a student’s experience of burnout. Regardless of other circumstances physics students have, being able to support and enhance their learning comes from understanding their experiences within the major.

One distinction helpful in determining the level of burnout students experience are the 5 stages of burnout identified in Ref. [8]. The first stage, the Honeymoon Phase, brings excitement, optimism, periods of productivity, and progress. The second stage, Onset of Stress Phase, is short-term periods of stress that bring a few physical and mental signs. The third stage, the Chronic Stress Phase, is described by persistent stress that can affect performance, negatively impact social relations, and bring feelings of pessimism. The fourth phase, the Burnout Phase, is where stress becomes a negative loop and does not relieve overnight. Usually, people are no longer able to function normally, and problems with work consume their lives. In the fifth stage, the Habitual Burnout phase, burnout becomes a daily part of life and can develop into serious depression or mental illness. Fatigue is dominant and work becomes hard to complete [9]. We will use these 5 stages allow us to understand the degree to which students experience burnout [8].

III. METHODS

In order to gain insight into the experiences of physics students, we developed a research study with two phases. The first phase focused on one-on-one interviews (with a short pre-interview survey) with undergraduate student volunteers. The interview protocol focused on open-ended questions de-
signed to elicit information on the student’s experiences with a high-pressure physics degree. The second phase of the research was informed by an analysis of the interview results from the first phase. Using these results we drafted a survey that could be distributed more broadly. Being one of the first to study burnout in PER, it is important to note that both the interview protocol and survey were informed primarily by results from occupational burnout studies and the Oldenburg Burnout Inventory.

Interviews in the first phase of research consisted of one-hour Zoom sessions. We interviewed seven students from two large research institutions at the start of the 2022 Fall semester. Additionally, the students were considered upperclassmen taking classes that involved multiple prerequisite physics classes to ensure they had been in the program for a minimum of two years since burnout is a long-term condition that builds over time. We recruited these students by email solicitation sent by professors teaching higher level physics core courses to distribute to their students.

The questions within the interview protocol focused on common symptoms of burnout and how students experience them. After a series of background questions about the student’s academic experience, we listed 18 symptoms of burnout and asked students to rank themselves on a 5-point Likert scale with respect to how often they experienced these symptoms, from always to never. To more accurately understand student experiences, we refrained from using the term burnout in either the pre-interview survey or the beginning of the interview protocol to avoid influencing students’ responses. Instead, burnout was not mentioned until the end of the interview study where we asked specific questions about the symptoms of burnout. We also asked students which of the 5 stages of burnout they most resonated with. In order to refrain from implying a diagnosis of students, we emphasized that we are not medical professionals but rather want to understand how physics students are feeling and experiencing their classes and coursework. Campus support resources were also provided to the students at the conclusion of the interview. We were able to categorize each student based on how high they ranked on the burnout scale and the symptoms they felt to search for correlations between the stages of burnout and extent of symptoms experienced.

To analyze the interview responses, we used qualitative coding to characterize students’ responses to common burnout conditions identified in existing literature. We categorized responses into \textit{a priori} codes that indicate burnout symptoms or feelings. We also identified emergent patterns we refer to as ‘anti-codes’ which were specifically present amongst students who had lower levels of burnout. These anti-codes identified strategies and characteristics that might explain why these students may experience less burnout.

Codes and anti-codes (see Table I) are defined such that there are no overlaps within the codes.

Based on the results of the interview study, we created a second survey to collect information on the experiences of a larger number of students. This survey was distributed, once again, through an email solicitation forwarded by professors teaching upper-division core courses. The survey was fully anonymous; however, to incentivize participation, students were provided an option to enter their email address through a separate interface to enter to win one of 5, $20 gift cards. Overall, 24 students responded to the survey. The survey focused specifically on identifying the symptoms and feelings of physics students. We narrowed the questions asked in the interview to focus on questions giving the largest amount of information on how students experience burnout. All of the closed-response questions asked in the interview study were included in the survey.

IV. RESULTS & DISCUSSION

To gain a sense of how many students were familiar with burnout as a concept, we asked participants to provide a definition of burnout in both the interviews and the survey. Every participant in the study was able to give a relevant and appropriate definition of burnout. The fact that students are (as expected) familiar with burnout lends credence to the concern surrounding students feeling burned out. Moreover, when participants were surveyed, 13 out of 31 students claimed they do not get enough rest always or frequently and 6 others said occasionally, indicating that many students are tired and may lack leisure time for self-care (which can lead to increased experience of burnout).

We focused our efforts largely on identifying how burnout presents and factors that increase or limit burnout. In this section, we will examine which symptoms are more frequent and how those relate to the different stages of burnout. We also investigated the effects of having a community vs. feeling isolated. Within the interview study, we found that students vary in burnout levels, but have similar patterns. Overall, 7 students were interviewed and 24 were surveyed for a total of 31 students. However, since some survey questions were optional or conditional, we have slightly smaller data sets for different parts of the survey. For example, if students did not identify with any of the 5 burnout phases they were not asked to report which symptoms they experienced.

The academic literature around burnout, along with our own experience, suggests that isolation can be both a cause and a symptom of burnout. Between the interviews and surveys, 7 students expressed they felt isolated or alone at college. However, we did not find any conclusive evidence that directly connected isolation (as a symptom or cause) with their experience of burnout, and, overall, these students experienced varying degrees of burnout, from the Honeymoon Phase to the Habitual Burnout Phase. We also asked students if they had friends in the classroom and had a sense of community. The remaining majority of students, 24 out of 31,

---

1 The full protocol can be accessed at https://www.dropbox.com/sh/vctjo6gserfbiyn/AAC6A0jAiKyUSDpwv-vUYyLta5Xdl=0
felt that they were not isolated and alone. Thus, in our data there was no correlation between isolation and the stages of burnout; however, the small \( N \) in our study limits the generalizability of this finding.

Within the interview study, one common trend shows that students generally do not ignore academic responsibilities, whether or not they experience burnout. Out of the seven interview participants, four students directly called academic responsibilities their number one responsibility, while the other three indirectly stated that they prioritize academia. Another commonality between the seven interviewees stated they felt supported by the classroom community regardless of if they had friends within the classroom.

The codes and anti-codes identified in the interview data are given in Table I. Codes indicating burnout are highly consistent with the literature. The only academic-specific code related to stress specifically around exams. Given the high-pressure nature of exams, this pattern is not surprising. The anti-codes, which were extracted from responses from students who were less burned out, show strategies or feelings that correlated with lower levels of burnout. These results suggest that there are certain actions, like careful time management and having support and passion, that students with lower levels of burnout incorporate. While it is not possible to determine, from these data, whether these strategies actively reduce a student’s experience of burnout, they do lay a useful foundation for future studies investigating strategies that might help students mitigate burnout.

Table II shows the frequency of the codes and anti-codes for each of the interview participants. Those with higher levels of burnout have a higher number of codes indicating burnout and very few anti-codes, e.g., Students 1 and 4. Those lower on the burnout scale still experience symptoms of burnout; however, they tend to have more anti-codes, e.g., Students 6 and 7.

In the surveys, our main focus was to identify the symptoms of burnout experienced by the 24 students surveyed. We asked students the same set of symptoms as the interview and the frequency in which they experienced these from always to never. Students were also asked to self-identify which of the five stages of burnout they most aligned with. Out of the 24 students who participated in the survey, 3 students selected none, 1 selected the honeymoon phase, 8 selected the onset of stress phase, 6 selected chronic stress, 1 selected the burnout stress phase and 5 selected the habitual burnout phase. We asked each student to identify how often they experienced the 18 symptoms of burnout seen in Table III and Table IV.

When examining the symptoms, we notice the physical symptoms are less present in students than the general set of mental and emotional symptoms. The most significant concerns are the 11 students who reported feeling ill, 12 students who reported stomach or bowel problems, and 12 students who experienced fatigue or neck pain at high rates of frequently or always. Some of the emotional and psychological symptoms also show high rates in the always or frequently categories, showing that a majority of students are fearful of failing and stressed or anxious. In general, students are physically healthy but have more physiological symptoms. Physics students are in high-stress situations and exams, and the data shows they experience varying symptoms as a result.

Table V reports the average number of symptoms per respondent for students in each of the 5 stages of burnout. In this table, we have categorized always and frequently together, as well as rarely and never to increase statistical power. Table V indicates that students in the Burnout or Habitual Burnout phases had more symptoms at higher frequencies. Students that were low on the burnout scale, showed

<table>
<thead>
<tr>
<th>Codes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Codes that Indicate Burnout</strong></td>
<td></td>
</tr>
<tr>
<td>Stress</td>
<td>Mental or emotional strain from the demand of responsibilities</td>
</tr>
<tr>
<td>Overwhelmed</td>
<td>Having too much school work or too much responsibility</td>
</tr>
<tr>
<td>Anxiety</td>
<td>Feelings of fear, dread or uneasiness</td>
</tr>
<tr>
<td>Exam stress</td>
<td>Stress correlation directly to exams or exam prep</td>
</tr>
<tr>
<td>Mental health</td>
<td>Comments regarding mental health issues</td>
</tr>
<tr>
<td>Isolation</td>
<td>Feelings of isolation from others</td>
</tr>
<tr>
<td>General unhappiness</td>
<td>Feelings that do not fall under other codes, but have negative connotations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Anti-codes that Limit or Lessen Burnout</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
</tr>
<tr>
<td>Social/Class support</td>
</tr>
<tr>
<td>Passion</td>
</tr>
<tr>
<td>Contentedness</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Burnout Phase</th>
<th>( N ) codes</th>
<th>( N ) anti-code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student 1</td>
<td>chronic stress (4)</td>
<td>18</td>
</tr>
<tr>
<td>Student 2</td>
<td>onset of stress (3)</td>
<td>2</td>
</tr>
<tr>
<td>Student 3</td>
<td>onset of stress (3)</td>
<td>5</td>
</tr>
<tr>
<td>Student 4</td>
<td>burnout (4)</td>
<td>8</td>
</tr>
<tr>
<td>Student 5</td>
<td>habitual burnout (5)</td>
<td>5</td>
</tr>
<tr>
<td>Student 6</td>
<td>honeymoon (1)</td>
<td>2</td>
</tr>
<tr>
<td>Student 7</td>
<td>honeymoon (1)</td>
<td>2</td>
</tr>
</tbody>
</table>

Table II. Frequency of Codes and Anti-codes experienced by the 7 interview participants and their corresponding stage of burnout.
fewer symptoms with lower degrees of frequency. Visually there is a clear indication for students falling into the none, honeymoon, and the onset of stress phases will have the majority of symptoms that rarely or never occur. This visual trend is supported using a chi-squared test (p < 0.00001) and a Cramer’s V = 0.3287, indicating a moderate association between the stages of burnout and the number of symptoms.

V. CONCLUSIONS & LIMITATIONS

To address the limited research on academic burnout, this study serves as a basis for future work within the field of physics education. We demonstrated links between the 5 phases of burnout and the frequency of burnout symptoms experienced by students. Our goal was not to identify how to solve or prevent burnout but, rather, to identify how it presents. We studied varying symptoms and have identified how often students experience these. The next step of this research is to determine how these symptoms affect academic performance through additional interviews and surveys. The end goal of this research is to develop resources to help students prevent the symptoms of burnout.

The results presented here have shown physics students have varying levels of burnout and those with higher levels are impacted emotionally, mentally, and physically, lending credence to the concern that many students are experiencing burnout. Regardless of the student’s level of burnout, interview trends showed that they are tired and drained, and most are not getting enough rest. Additionally, almost all students were fearful of failing classes. Notably, selection effects regarding who typically responds to requests for survey suggest our results should be interpreted as a lower bound on the frequency of these symptoms amongst the broader physics student population. Recognizing these patterns serve as a useful step in aiding physics students and gaining better academic results. One important aspect missing from this research consists of the students with no burnout. Only 3 of our students identified as having none of the 5 stages; however, we did not require these students to answer the symptoms questions. In future work, it would be valuable to ask students who don’t self-identify as having burnout what their symptoms are as well for a base comparison. Expanded future work would also address another limitation of the current work - small N and a student population restricted to large, research-intensive institutions. However, despite these limitations, this research gives an insight into the presentation of burnout in physics students and serves as a basis to build from.

ACKNOWLEDGMENTS

This work was supported by the University of Colorado Boulder and the CU PER team.


Testing for over- and under-dispersion in physics degree outcomes.

Astra Sword

School of Physical Sciences, The Open University, Milton Keynes, England, United Kingdom, MK7 6AA

As the scale of quantitative data available to physics education researchers grows, it is imperative that we critically assess how well the assumptions behind "standard" statistical methods apply in our field. In the present work, I give a background on a common statistical assumption used to analyse proportion data, the binomial assumption; I discuss scenarios in which this assumption may break down in the context of education research; and test this assumption using a large population-level data set. This data set comprises academic outcomes (rate of 'good degrees') for all undergraduate physics degree programs that ran in the UK across the 2012/13–2018/19 period (26,960 students across 79 programs). I estimate dispersion parameters and their significance for each program in the data set and discuss the implications of the results for analysing proportion data in physics education research.
I. INTRODUCTION

As quantitative physics education research increasingly makes use of larger data sets [1] there is a great potential for producing broad, wide reaching results that can help inform policy and practice on a national, and even international, level. However, with the scale of the data, the impact of errors in analysis also grows. As such, special care must be taken with the application and interpretation of the statistical methods used to analyse these data--with a critical view to understanding where the (often unspoken) assumptions that underlie these standard techniques break down.

This is particularly true for quantitative diversity, equity, and inclusion (DEI) research, where the principal research subjects of interest are people from socially disadvantaged—and consequently, often vulnerable—demographic groups. Within PER, there is a rich and growing literature addressing the breakdown of conceptual and statistical assumptions in PER-related DEI research addressing topics like the problems with using a binary framework for gender [2], omitted variable bias in the analysis of demographic gaps [3], and general issues of causal inference [4].

In this paper I contribute to this wider body of work by using a large national data set containing degree outcomes for UK physics graduates to test for violations of a particularly ubiquitous statistical assumption: the binomial assumption. Along with its generalised (multinomial) form, this assumption underpins a number of techniques used to analyse these data--with a critical view to understanding where the (often unspoken) assumptions that underlie these standard techniques break down.

Within PER, there is a rich and growing literature addressing the breakdown of conceptual and statistical assumptions in PER-related DEI research addressing topics like the problems with using a binary framework for gender [2], omitted variable bias in the analysis of demographic gaps [3], and general issues of causal inference [4].

In this paper I contribute to this wider body of work by using a large national data set containing degree outcomes for UK physics graduates to test for violations of a particularly ubiquitous statistical assumption: the binomial assumption. Along with its generalised (multinomial) form, this assumption underpins a number of techniques used to understand differences between demographic groups for categorical outcomes. Examples include the $\chi^2$ goodness of fit and independence tests; the Cochran-Mantel-Haenszel procedure, often used in differential item functioning analysis; as well as various forms of regression such as log-linear analysis and logistic regression [5].

I present:

- (Section II.) A background on the binomial assumption, how it might break down in the context of academic success, and what circumstances might lead to overdispersion and underdispersion respectively.
- (Section III. & IV.) An analysis of dispersion parameters and their statistical significance for all physics degree programs in the UK using a simulation-based approach from the R library DHARMa [6].
- (Section V.) A discussion of the possible implications of the results in the context of UK higher education.

The research question for this study is: "For each degree program in the data, does the observed variance in the good degree rate match the expected variance under the binomial distribution?" See Section II. & III. for definitions.

In addition to making a general contribution to the literature, this work is part of a wider project using the aforementioned large national data set to understand what the demographic gaps in physics are, their causes, and what role physics-as-a-subject plays in their evolution. It is hoped that the results of this analysis will help inform similar projects in the future.

II. BACKGROUND

A. The Binomial Distribution

The binomial assumption is the assumption that a collection of binary response variables, for example $Y_i \in \{\text{Failure}, \text{Success}\}$, are Bernoulli variables that are both (a) identically distributed and (b) independent.

$$Y_i \sim \text{Bernoulli}(\pi)$$

In other words, $Y_i = 1$ with a probability $\pi$ and $Y_i = 0$ with probability of $(1 - \pi)$. This is equivalent to declaring that your response variable is generated by a process equivalent to flipping a weighted coin that has probability $\pi$ of getting heads for each observation $i$.

In practice, we usually observe binary response variables (such as whether a student has passed a particular course) in clusters (i.e. cohorts). If we observe $K_j$ successes per cluster of size $m_j$, we can say $K_j$ is independently identically distributed according to:

$$K_j \sim \text{Binomial}(m_j, \pi)$$

Both forms for the data (1) and (2) are strictly equivalent; though usually (2) tends to be the representation used when dealing with proportion data, such as the fraction of students passing a course. In this case, the expected pass rate for a cluster is $E(Y_i) = \pi$ with a variance of $\text{Var}(Y_i) = \pi(1 - \pi)$. Alternatively we could express these as the mean and variance of $K_j$ instead; $E(K_j) = m_j \pi$ and $\text{Var}(K_j) = m_j \pi (1 - \pi)$.

This brings us to the crux of the issue: the variance of binomially distributed data is uniquely determined by the mean. (Contrast this with normally distributed data, where the mean and variance can vary freely.) Having only one free parameter results in the fit between the binomial distribution and the real data being more fragile.

B. Relaxing the Binomial Assumptions

As noted, modeling the pass rate of a cohort of students as binomial is equivalent to declaring (a) that each student has exactly the same chance of passing and (b) that success of each student is independent of every other student. Both these seem unlikely, consider these hypothetical motivating arguments for why these assumptions may not apply:

(a) Diverse chances of success. We expect a student’s chance of success to depend on what they already know—experience suggests this is unlikely to be the same for every student, and so chance of success should be different for different students too.

(b) Correlated outcomes. Students do not study in isolation—in fact, they are typically encouraged to form study groups for working on problems. For assessments where students can share answers, success is likely to be correlated within each study group.
While these two arguments may not be universally applicable to every classroom, their plausibility suggests that we should expect student outcomes to not meet the assumptions of the binomial distribution at least some of the time.

1. **Diverse chances of success**

When the assumption of identical distribution is not met, the extra uncertainty in the chance of success $\pi$ leads to an **overdispersed** distribution [7], that is to say a distribution with larger variance than would be expected from a binomial model with the same mean value of $\pi$. In practice, this can be modeled by specifying an underlying distribution for $\pi$. The most common choice is the beta distribution which results in using a beta-binomial distribution to model the data [8]. Figure 1 shows an example overdispersion due to beta-distributed $\pi$. Note that allowing $\pi$ to vary always leads to a distribution that is overdispersed relative to the binomial [7].

2. **Correlated outcomes**

When the assumption of independence between outcomes is not met, there are two main possibilities. If outcomes are positively correlated—such as when students working together on a problem submit similar answers—this also leads to the data being overdispersed relative to the binomial [7]. However, if outcomes are negatively correlated—one student’s success makes other students less likely to succeed—then the data will be **underdispersed**, that is exhibit less variation than the corresponding binomial distribution [7].

What could cause student outcomes to be negatively correlated with one another? While this is not something I have seen discussed in the wider literature in the context of underdispersion, I suggest at least two plausible mechanisms:

- **Shared, finite resources.** While students typically have access to the same lectures and materials, some academic resources are necessarily finite and must be shared with other students. Possible examples include informal one-to-one tuition from instructors; places on enrichment programs; or access to non-academic resources that may contribute to success such as affordable housing close to campus.

- **Post-hoc marking of assessments.** This is the adjustment of student scores and adjustment of grade boundaries for an assessment after it has taken place. Specific post-hoc marking practices are diverse, but may include adjusting the pass mark so the expected number of students pass [9]. As a consequence, each student that fails the assessment increases the chance other students will have their mark adjusted past the pass threshold post-hoc, leading to negatively correlated outcomes.

C. **Implications for Statistical Analysis**

It is important to note that under- or overdispersion of a outcome variable across any particular groupings within a data set is not necessarily a statistical issue in and of itself. Statistical techniques that rely on the binomial or multinomial assumptions, like those listed in the introduction, only require that those assumptions hold for the outcome variable conditioned on the regressors included in the model [10].

In practice, it is often impractical or impossible to take into account all the factors that may impact student outcomes. The extent of non-binomial variation in a data set can be summarised using the dispersion parameter :

$$\phi = \frac{\sigma^2_{\text{obs}}}{\sigma^2_{\text{exp}}}$$

(3)

Where $\sigma^2_{\text{obs}}$ is the variance observed in the data set and $\sigma^2_{\text{exp}}$ is the variance expected under a binomial assumption. If overdispersion ($\phi > 1$) is present, models using the binomial assumption will more likely to produce false-positives, while for underdispersion ($\phi < 1$), the false-negative rate will be higher. In both cases the true confidence interval (CI) can be estimated by multiplying the naïve CI obtained under the binomial assumption by $\sqrt{\phi}$ [8].

D. **Further Reading**

Several standard texts address overdispersion, albeit briefly. See [11] for a brief overview. For a more technical treatment [8][10] are good choices, with the former being focused exclusively on categorical statistics and the latter having broader coverage of statistical models. Where underdispersion is noted in these and similar texts, it is usually just to mention that it is rare. However, [12] does have a chapter introducing both over and underdispersion, as well as a chapter with material on under-dispersed Poisson models.
III. METHODS

If Section II. establishes plausible scenarios for non-binomial dispersion the question remains: is there empirically relevant over- or under-dispersion in academic outcomes in physics? To answer this question, I estimated the root dispersion parameters and their statistical significance for a large data set comprising all graduates of accredited physics degree programs that ran in the UK across the full 7 years spanning 2012/13 and 2018/19 (26,960 students; 42 institutions).

A. Context

In the UK, undergraduate degree programs typically focus on a single subject, chosen by the student at the start of their studies. On graduation, students are awarded a bachelor’s degree (BSc) with a classification determined by their academic performance. Of these, the highest are the first class (1st) and upper second (2:1) which are roughly equivalent to a 3.7+ GPA and a 3.3+ GPA respectively [13]. The percentage of graduates on a particular program or at an institution achieving a 1st or a 2:1 is known as the *good degree rate* and is used across the sector to measure academic success.

At most UK institutions delivering physics programs, high performing students often have the option to extend their program by a year to gain an "enhanced" degree, graduating with a master’s (MSc) rather than a "non-enhanced" bachelor’s level qualification. Broadly these two programs will be the same for the first three years at any particular institution, with the students being in a common cohort and having the opportunity to transfer from and to the enhanced program—though the exact details vary by institution.

Both enhanced and non-enhanced physics programs in the UK undergo accreditation by the Institute of Physics (IOP), which requires accredited programs to teach a standard set of topics [14].

B. Data Source

The data in this study is a subset of a larger data set procured from the Higher Education Statistical Agency in the UK (HESA) for the purposes of studying equity in physics degree outcomes. This larger data set includes all students that have studied first-degrees in the UK between 2012/13 and 2019/20, not just physics students, and comprises approximately 4 million students in total.

Due to the sensitivity of this data, my reporting is subject to a number of stipulations [15] including:

(a) Counts of students are rounded to the nearest 5.
(b) Individual higher education providers must not be identifiable.

A consequence of (b) is that I am unable to report the exact size of any of the specific physics programs investigated here as that information could be used to identify them.

C. Identification of Physics Degree Programs

I identified physics degree programs as those that met the following criteria:

1. The degree program is accredited by the IOP [14].
2. The degree program has "physics" in the title (i.e., is not a natural science degree).
3. The program’s accreditation is for the degree program as a whole, rather than just for a subset of students on the program that take specific modules.

I considered all programs that met these criteria at a particular institution and that led to the same level of qualification (i.e., enhanced vs non-enhanced) to belong to the same physics program.

D. Statistical Inference

While the dispersion parameter (3) is the standard way of measuring non-binomial dispersion, to aid interpretation I calculated the root dispersion parameter instead:

\[ \sqrt{\phi} = \sqrt{\frac{\sigma_{\text{obs}}^2}{\sigma_{\text{exp}}^2}} \approx \frac{\text{True CI}}{\text{Naïve CI}} \]  

Where \( \sigma_{\text{obs}}^2 \) was the observed year-on-year variance calculated across the 7 years in the sample, and \( \sigma_{\text{exp}}^2 \) was the variance expected under a binomial assumption. In theory, for a program of fixed size each year, the expected variance in the good degree rate \( \pi \) would simply be \( \pi(1 - \pi) \) and thus \( \sigma_{\text{exp}}^2 \) could be calculated analytically. However, as the number of students graduating in any particular cohort varies from year to year, a different approach was needed. For this analysis, I used the `testDispersion` function of the R library DHARMa [6] to estimate the distribution of expected variance via simulation. I used 1,000,000 iterations to estimate the distribution under the null hypothesis for each program.

Post-hoc Analysis. Given that this analysis involved a large number of significance tests, I used the `statmodels` [16] implementation of the Benjamin-Hochberg procedure to control the false discovery rate at 5% [17].

IV. RESULTS

a. Descriptive statistics. Across the 42 intuitions in the data set, there were 42 non-enhanced (BSc) and 37 enhanced (MSc) degree programs—this is in line with expectations as most institutions deliver both types. The average cohort size per year varied considerably for both non-enhanced (\( M = 45, SD = 20 \)) and enhanced programs (\( M = 50, SD = 40 \)), while the average good degree rate per year across the period varied moderately for both non-enhanced (\( M = 57\%, SD = 15\% \)) and enhanced programs (\( M = 89\%, SD = 10\% \)). The total number of graduates of non-enhanced and enhanced programs were similar, 13,465 and 13,495 graduates respectively.
b. Root dispersion parameters. Figure 2 illustrates the distribution of estimated root dispersion parameters for both non-enhanced (M = 1.10, SD = 0.29) and enhanced degrees (M = 1.02, SD = 0.12), with those that were significant at the α = 0.05 level highlighted.

c. Significance. Prior to adjusting the p-values via the Benjamin-Hotchberg procedure, 16 programs showed significant non-binomial dispersion (Table I). Controlling the false discovery rate, only programs A and G were found to have significant non-binomial dispersion at the α = 0.05 level; though a further 4 were ‘near misses’ at p < 0.051.

VI. DISCUSSION

In the context of UK based education research, this data demonstrates the existence of physics programs with non-binomially distributed good degree rates: namely programs A and G showing clear statistical significance; with programs E, J, L and K being very close to significance. The root dispersion parameters cover a wide range, from 0.41 (Program J) to 1.9 (Program A). Together with Equation 4, these parameters suggest the true confidence intervals for these programs range from under half through to nearly twice the confidence intervals that would be calculated under the binomial assumption. These large departures from the binomial assumption have clear empirical relevance for how differences in good degree rate should be interpreted between these programs.

The presence of underdispersed (√φ < 1) programs is particularly interesting given that underdispersion arises from the anti-correlation of outcomes between students. This may reveal that for these programs the mechanisms suggested for the anti-correlation in Section II. (e.g., resource scarcity, posthoc marking) have a dominant presence in these programs compared to the mechanisms that lead to overdispersion (e.g., diverse success rates, students working together). Similarly, the stark contrast in the spread of root dispersion parameters between program types may be reflective of the higher entry requirements for enhanced programs, which leads to a narrower range of individual student success rates—thus reducing the extent of any overdispersion.

For PER as a field, the presence of statistically significant and empirically relevant non-binomial dispersion in this data set suggests that the validity of the binomial assumption cannot be taken for granted. I propose three recommendations for PER researchers to account for non-binomial dispersion going forward:

- **Interpret binomial tests with caution.** Consider reporting how much non-binomial dispersion would be needed to alter the interpretation of any associated p-values or confidence intervals.
- **Use dispersion-aware models.** Quasi-likelihood models can account for both over- and underdispersion. If dealing with overdispersion only, a parametric or mixed effects approach may be better. [8] [10]
- **Measure it!** Non-binomial dispersion is a feature, not a bug. Its size and direction is a clue about what is happening in the classroom. The dispersion parameter can be quickly estimated from goodness-of-fit statistics included the output of statistical packages (see [11]) but a simulation based approach is more robust [6].

VI. ACKNOWLEDGEMENTS

Special thanks to Sally Jordan, Annika Lohstroh, Rachel Hilliam, and Mark Jones for their encouragement and feedback on the draft of this paper. This project was funded by The Ogden Trust and The Open University.


Characterizing the complexities of experimental decision making in an introductory lab practical

Hamideh Talafian and Katie Ansell
Department of Physics, University of Illinois Urbana-Champaign, 1110 W. Green St., Urbana, Illinois, 61801, USA

This study characterizes the complexities of decision-making in an investigation-style practical in two case study groups with seven students in an introductory physics course. The students were trying to test a claim in a hypothetical situation about the effects of drag on the time it takes for a cotton ball to fall. Using two videos as case studies, the study focuses on the complexities of students’ experimental decision-making centered around three metacognitive skills: 1) planning prior to investigation, 2) monitoring while doing the experimentation, and 3) evaluating the results. We characterize the complexities of experimental decision-making by creating visuals that show the occurrences of each code in each stage of decision-making. We concluded that the group with more complex approaches to experimental decision-making, characterized by back and forth between activating multiple metacognitive skills, were able to connect the topic to other physics concepts, have more effective conversations around discrepancies of data in multiple trials, and loop back the results of the experiments to the original claim. However, the other group showed a more linear effort, lacking planning prior to investigation and evaluating the results. We argue the complexities of experimental decision-making can be an indicator of the presence or absence of some of the metacognitive skills in lab activities, but further studies need to confirm this finding before making a generalizable comment.
Investigation-style laboratories in introductory physics courses are spaces with a unique potential to develop students' fluency with scientific skills, experimental processes, and their abilities to generate and coordinate ideas and solutions [1]. These labs (i.e., in contrast with step-by-step verification labs) are often paired with the term ‘inquiry’ and claims of teaching critical thinking as a metacognitive skill that can promote ‘authenticity’ in the lab space [2–8]. Such labs differ from concept-focused instruction in that they prioritize experimental processes and student decision-making over results [9]. In investigation-style labs, specific learning objectives can be embedded into experiments such that they arise naturally and must be confronted before the experiment is complete [10]. Thus, students are provided with scaffolded opportunities to practice and develop ‘good thinking’: Bailin [11] argues that critical thinking is not best defined by lists of processes or skills, but rather good thinking in the norms of a particular context. Research on introductory labs often assesses ‘good thinking’ in terms of the execution of metacognitive skills, alignment of attitudes with experts, or alignment of reasoning processes with experts [12, 13]. Although these triangulate on components of ‘good thinking,’ they fall short of giving a clear picture of what decision-making in lab experiments looks like when students activate various metacognitive skills.

Previous scholarship recommends researching students’ metacognitive activity, especially during lab and experiment activities [14]. In this study, we shed some light on the nuances of decision-making as connected to three prevalent metacognitive skills by looking at two seemingly similar videos in which the participants showed engagement and collaboration as a group. By focusing on those complexities, we aim to explore the connection between metacognitive skills and the complexities of decision-making in experiments.

I. BACKGROUND

In investigation-style labs, students who collaborate in a group solve problems for which the path to the solution is not necessarily obvious or linear. Such labs require activating multiple metacognitive skills such as critically thinking about the situation, reflecting on collected data, and evaluating results. However, despite knowing the critical role of metacognition in learning [15], metacognitive skills are challenging to develop and measure in physics problem solving in general and in labs in particular [16]. Work on this topic can be categorized into three types: 1) studies that focus on specific metacognitive skills in labs and learning outcomes (e.g., troubleshooting and their impact on problem solving [17], mathematical problem solving [18], and reading comprehension of physics texts [19]), 2) studies that focus on improving students’ attitudes toward the nature of science by encouraging metacognitive skills that help them control their own learning such as reflection, opportunities for providing feedback and modification [20–22], 3) studies that focus on collaborative problem solving and developing metacognitive skills, e.g., [23–25]. Tasks with high levels of operationalization and complex problem-solving have been shown to support the development of metacognitive skills [26]. In the context of investigation-style physics labs, which align with tasks that elicit metacognition, the complexity of pathways emerges as an important factor worthy of investigation.

Using the lens of descriptive decision theory [27] which considers how individuals make decisions in situations with multiple possibilities and/or uncertainty, we embrace the diversity of experimental decision-making among students as a complex process that needs to be further unpacked to encourage new pedagogical practices in teaching physics labs. Descriptive decision-making differs from normative decision-making in that we focus on describing how groups navigate decisions without defining optimal behavior.

II. METHODS

This work focuses on two videos of student groups in an introductory Calculus-based Mechanics course at the University of Illinois doing an end-of-semester lab practical experiment. We use a qualitative approach to characterize the discussion topics and actions of each group, presenting both overall snapshots of each group’s behavior and specific episodes that highlight how complexity and metacognition appear.

A. The lab practical experiment

At the end of the semester, students were given an in-class lab practical exam to assess their lab competencies and skills. The exam had two components: an individual written component focusing on proposing experiment design as well as data analysis and conclusion-making, and a group experiment component where students designed and performed an experiment to test a claim given in the task prompt and wrote a brief, informal report of their investigation. Students were given an hour to complete the group experiment, but most were done in 20–40 minutes.

In the task prompt, a hypothetical peer (“Jamie”) proposed that the drop height of a cotton ball was sufficient information to accurately predict the cotton ball’s fall time. Students were prompted to write an explanation of the reasoning for Jamie’s claim, and to design, do, and report on their own experiment that tested Jamie’s claim. A supply table in the classroom contained materials available to students for the activity, including cotton balls, meter sticks, stopwatches, and digital scales. Students were also told that they could use any resources including the Internet, but that their instructor could only answer questions to clarify the wording of the prompt.

Most groups were able to use Jamie’s reasoning to problematize the claim and subsequently design an experiment and take measurements. Many groups utilized an experi-
ment design that involved dropping cotton balls from various heights and measuring fall times with the stopwatch. Because the fall times of the cotton ball were on a comparable scale to human reaction times, students encountered difficulty collecting data and had to adapt to deal with anomalous results. All groups were eventually able to collect what they deemed sufficient data to consider the task complete.

B. Video analysis methods

Sixteen groups were recorded during the group experiment using video cameras mounted nearby and voice recorders on the groups’ tables. Half of the groups had experienced a reformed lab format during the semester. Videos from two groups from the reformed format were identified for analysis in this work by the contrast of their complexity, but the relative similarity of their other characteristics: Both groups were taught by the same instructor, demonstrated high levels of group coherence, were similarly perceived as efficient at a surface level, and spent similar total amounts of time on the lab practical experiment. These groups are generally representative of other groups in the class in terms of composition and ability, but represent two extremes on the spectrum of complexity. Group 1, who we identify as the ‘complex’ group, consisted of three students (two men and one woman), who spent twenty-two minutes working on their experiment. Group 2, who we identify as the ‘linear’ group, consisted of four students (three men and one woman), who spent twenty-nine minutes working on their experiment. These times were typical for most groups in the class.

The authors generated a two-level coding scheme for student groups working on the lab practical. At the first level, the researchers used a grounded approach to establish areas of experimental decision-making, i.e., events or stages in which groups were deciding what to do. Videos from several groups were used to generate independent codebooks (see [28] for an expanded initial list), which were resolved by the authors into a single list focusing on decision-making discourse. The full list of decision-making codes is presented in the first column of Table I. At the second level, the researchers grouped the decision-making codes into three categories for metacognition: planning, monitoring, and evaluating.

Two researchers coded a transcript from another video and reached a fair agreement (fifty percent of the time) which resulted in refining the coding scheme and assigning descriptive notes to each code. The discrepancies were around the “strategizing” and “discussing design” codes which were resolved by discussing the episodes in which they occurred. For the two videos discussed in this study, both researchers reached full agreement in assigning codes to specific segments. Video data was segmented for assigning codes based on the semantic nature of the episode (i.e., smallest thematic pieces of the discussion) rather than timestamps. For this preliminary analysis, one researcher coded both videos twice within a time span of two months. All data were analyzed in MAXQDA 2022, and the generated visual shown in subsequent sections is a depiction of the coded segments called a document portrait. The document portrait provides a qualitative overview to identify patterns. Coded segments of the videos were in different sizes ranging from a few seconds to a few minutes. We found the document portrait a powerful visual that weights the coded segments based on the duration of the video. In these visuals, fewer dots represent shorter segments of the video, and more dots represent longer video segments. Segments with multiple codes are presented in sequence (i.e., if a segment is assigned two codes, a set of dots would be shown in the first color, then repeated in the second color).

III. RESULTS

Preliminary results of our video analysis showed considerable differences between the groups’ experimental decision-making codes. Table I shows the absence and presence of each code for each group associated with metacognitive skills. To better clarify the differences between each group, we narrate the collaboration of two groups’ entire session accompanied by document portraits (Fig. 1) visualizing coded segments in the following sections.

A. The Complex Group

The complex group began their conversation by reviewing the prompt and deciding that Jamie is wrong because they only considered height and gravity to understand the situation. Student B shows her agreement with student A who first expressed his disagreement with Jamie’s claim and called Jamie “an idiot” for not mentioning the role of drag. The group then spent one minute and twenty-two seconds (coded as light blue) discussing the assumptions that supports what

<table>
<thead>
<tr>
<th>Code</th>
<th>Group 1 (complex)</th>
<th>Group 2 (linear)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predicting outcome</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Theorizing</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Discussing prompt</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Discussing design</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Proposing design details</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Monitoring</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strategizing</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Restategizing</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Evaluating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Making sense of results</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Deciding about the claim</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Disagreement/confusion</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
they mentioned earlier about Jamie being wrong. They reviewed physics concepts related to kinematics to understand the reasoning behind Jamie’s claim by maintaining their original stance. As the conversation progressed, the group brainstormed possible ways of executing the experiment. At this point, the conversation became complex as the group discussed the design procedure, strategized about how to do the experiment by one member proposed a potential way of doing it which culminated in restating that “Jamie is wrong” which looped back to reading the prompt one more time and rediscussing the design procedures.

One important finding here is that the group does not follow a linear model of doing a lab by reviewing the prompt, discussing the design procedure, proposing a design procedure, and executing the experiment. There are lots of evident back and forth in this episode that we can unpack by looking at the section highlighted with a black box in Fig. 1a and the corresponding student discussion, shown with assigned codes and accompanied possible metacognitive skills that are being activated in Table II.

In this highlighted episode, discussing the design procedure followed up by strategizing and proposing design details leads this group to revisit their decision-making. This in turn leads them to revisit the prompt (orange in Figure 1). Row 2 in Table II shows the associated conversation. We also note the appearance of the blue code (from row 3) or “making a decision about the claim” during this planning stage as the group thinks ahead to the end goal of their experiment. Based on Figure 1a, we can say this is the second time the students are getting back to their original decision about Jamie being wrong.

After proposing and discussing the design procedures, the group realized they actually needed to collect data. An experiment design comes together quickly, with planning and regulation occurring as one of the group members assures another that “It’ll be really easy” to make measurements and test the claim, and a later reality check that although the group planned nine trials they would, in fact, go very quickly. As the group transitions to collecting data, they take about 30 seconds to assign roles (dropping the cotton ball, operating the stopwatch, and recording results) and do a test drop. Over the next 2–3 minutes they collect data for their three trials at each of the three heights; the student operating the stopwatch sometimes asks to redo a trial immediately after hitting stop and reading the time. The group goes back and forth between discussing the design procedure (doing different trials) and proposing the details of the design. For instance, at the high-

<table>
<thead>
<tr>
<th>Transcript</th>
<th>Code(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student A:</strong> We are just dropping it from different heights and calculating the air resistance?</td>
<td>Discuss design, Strategize, Propose design</td>
</tr>
<tr>
<td><strong>A:</strong> Isn’t our conclusion that he is wrong...doesn’t like there is a drag</td>
<td>Decide about claim</td>
</tr>
<tr>
<td><strong>Students A and C:</strong> Let’s just do it</td>
<td>Discuss prompt</td>
</tr>
<tr>
<td><strong>B:</strong> Wait do we have to do it?</td>
<td></td>
</tr>
<tr>
<td><strong>A:</strong> Yes, [reads the prompt] design an experiment...</td>
<td></td>
</tr>
<tr>
<td><strong>B:</strong> Do we have to perform it?</td>
<td><strong>A:</strong> Yeah <strong>B:</strong> Ughh</td>
</tr>
<tr>
<td><strong>A:</strong> How are we gonna test it?</td>
<td>Discuss design</td>
</tr>
<tr>
<td><strong>B:</strong> Just different heights, then we drop it.</td>
<td></td>
</tr>
</tbody>
</table>

FIG. 1: Document portraits indicating the order of coded segments as each group progressed through the laboratory practical exam. Silent periods are not represented in these images.
est height some group members become concerned by large time measurement variations for the 2 m height, and notice that the cotton ball is falling in an “arc.” They decide to try to avoid an arc shape drop and repeat the experiment. They repeat measurements until they have satisfactory values for all trials and conclude the collection of data. After this point they transition to a more linear, efficient stage to make calculations and establish conclusions about the results, “Jamie is wrong because they haven’t considered drag,” which has been previously coded at the beginning of their conversation. There was some confusion about Jamie’s correctness, but the team swiftly resolved it, reaching a consensus that Jamie was indeed wrong. The disagreement/confusion code was not limited to specific results and could appear anywhere. However, its presence exclusively in the complex group strengthened the final conclusion, reflecting everyone’s agreement.

B. Group 2: Linear Decision-Making Processes

This group demonstrates an example of taking a mostly linear approach in decision-making, in which each step leads to the next one and we see almost no trace of other codes between each episode of decision-making. As shown in Fig. 1b, the group begins by reading and discussing the prompt and reviewing the assumptions, showing their only, short-lived entry into complexity around the “discuss prompt” codes. This leads to a quick proposition of design procedure by one group member saying, “Should we just drop them?” not referring to a specific object of the procedure. The conversation resumes when the group reviews some assumptions related to the prompt which requires rediscussing the prompt to understand what it wants and reviewing the assumptions: “So we are discussing the reasoning behind Jamie’s claim, which is he is only using the kinematics equation based on the fact that there is only one force acting on the cotton ball, which is acceleration due to gravity.” Here, the students enter the next phase, discussing the design procedure. In contrast to the complex group, whose conversation we highlighted in the previous section, the bulk of this group’s equivalent planning conversation was, “A: So the experiment is dropping the cotton ball. B: yeah.” This discussion around procedure disconnects once when they strategize about drop height and then again continues for a few more minutes. After making small tweaks to the procedure during data collection, the students try to make sense of their results by plugging their numbers into the equation and checking a few things between each other, “We don’t really need the mass, right?” The group’s conversation never enters an interpretation mode and from the video, it was not clear what they decide about Jamie’s claim.

One important finding here is the absence of some codes such as “predicting the outcome” which could happen earlier in the conversation and “making a decision about the claim” which happened nowhere in the video. We also note the uniformity of each code once they enter each phase. For instance, while the group discusses the design we observe little to no back and forth between other codes. These students followed a logical pattern which may be considered to be more efficient. However, the low level of complexity may also reveal a lack of planning prior to investigation and evaluating the results which are important in this context for conducting the lab at a more sophisticated level.

IV. DISCUSSION AND CONCLUSIONS

The two case studies provide a juxtaposition between groups that engage in highly complex discussion with several metacognitive elements, and groups that are organized and efficient, but lack the complexity of the former. This preliminary complexity-metacognition mapping may offer a framework to identify or characterize ‘good thinking’ in the context of physics laboratory group work. We argue that these complexities — characterized by back and forth between different types of experimental decision-making — are important in activating metacognitive skills. Previous work has depicted the linearity of experimental decision-making by depicting the logical sequence of experimental decision-making (e.g., the ISLE cycle in [29]). Our preliminary results painted a more complex picture of experimental decision-making when students were executing physics labs. These complexities were associated with more effective conversation around discrepancies of data, connection to the topic of drag, and better collaboration among group members which were not shown in the analysis of the group who showed a linear approach.

We have considered groups on two extremes of linearity and complexity to explore a way to understand how groups progress through experiments. It is noteworthy that the linearity or complexity of experimental decision-making is not positive or negative per se. What matters is the extent to which it changes the behavior and direction of the experiment as previous work suggests [16]. From the metacognitive standpoint, previous work has shown that complexity of decision-making is an expert trait which keeps experts on track in problem-solving while lack of those may result in poor performance [30]. Hence, future work should take a similar approach to identify possible patterns that could make the connection between metacognitive skills and instant decisions clearer. These preliminary findings will also inform future studies for developing teaching strategies that could encourage the use of these skills in lab design in physics.

ACKNOWLEDGMENTS

We would like to thank the members of the Physics Education Research group at the University of Illinois Urbana-Champaign for providing feedback, particularly Tim Stelzer and Eric Kuo.
Using Community Cultural Wealth to Understand Experiences in Physics Bridge Programs

Jenna P. Tempkin
Department of Physics, Lafayette College, 111 Quad Drive, Easton, PA, 18042

Geraldine L. Cochran
Department of Physics, The Ohio State University, 191 West Woodruff Ave, Columbus, OH 43210

Téa Boone
Department of Physics and Environmental Science, St. Mary's University, 1 Camino Santa Maria, San Antonio, TX, 78228

Many conceptual and theoretical frameworks in education research tend to value social and cultural capital from the most privileged groups and focus on what people from marginalized groups are “lacking.” This kind of research often employs a deficit model of understanding the experiences of people marginalized in education. To fully understand the academic experiences of students from marginalized groups it is crucial to focus on the capital these groups use to overcome challenges. The Community Cultural Wealth (CCW) framework values resources Communities of Color have to help them through various systems such as education. The goal of this project is to understand which forms of CCW students utilize when deciding to pursue graduate education in physics and when they face challenges while in graduate school. To do so, we conducted semi-structured interviews with participants in physics bridge programs from 4 different institutions, and used qualitative coding to identify examples of each of the 6 types of CCW at three different time frames: pre-college, deciding to pursue graduate school, and during graduate school. Our analysis of the data showed that students used familial and aspirational capital during their pre-college experience and navigational and social capital when deciding to apply for graduate programs. Students tended to utilize familial, social, and navigational capital once in their program. We also highlight the impact of family and socioeconomic and ethnic background on experience of students in the APS Bridge Program.
I. INTRODUCTION

Recent physics education research studies reveal a number of challenges that physics graduate students face. For example, Barthelemy, McCormick, and Henderson (2016) [1] documented graduate student experiences of sexism and gendered microaggressions and suggested training and policies to combat the gendered experiences faced by graduate students in physics and astronomy. In their investigation of physics identity development, Hyater-Adams, et al., (2019) [2] found that Black physicists have racialized experiences in physics that influence identity development and experience within the field. They suggest that we interrogate material resources, such as bridge programs, to understand if/how these programs provide students with positive recognition that can influence their physics identity development. Additionally, Sachmpazidi and Henderson (2021) [3] found that in general, physics graduate students experience limited social and academic integration. Through the American Physical Society Bridge Program (APSBP) however, they found that departments can create structures that support student social and academic integration. We cite these three studies as examples of the work being done in physics education research to identify and understand the challenges that physics graduate students face and also provide solutions for facing such challenges. In this study we aim to contribute to this growing body of work by understanding the resources that graduate physics students in physics bridge programs use to overcome the challenges they face in graduate school and as part of their decision to pursue a doctoral education in physics.

Physics education researchers have advocated for research that factors in the demographics of the participants [4] and that utilizes appropriate framing when studying marginalized populations in physics [5]. Thus, we developed our conceptual framework on Community Cultural Wealth (CCW), which utilizes Critical Race Theory to challenge "traditional interpretations of cultural capital" (Yosso, 2005). More specifically, we focus on the experiences of graduate students from minoritized and marginalized ethnic/racial groups using a framework that values the resources that these students bring to academic environments.

A. The American Physical Society Bridge Program

This study focuses on participants of the American Physical Society Bridge Program. The APSBP is designed to increase the amount of physics PhDs awarded to Black, Latinx, and Indigenous students by providing structured support and resources for one’s transition into graduate school. The APSBP collects applications from candidates after the typical mid-April cut-off and sends them to participating institutions, allowing the institution to select candidates that best fit their particular program. Once admitted into the program, the APSBP provides students close support by boosting social integration and mentoring of incoming students. Sachmpazidi and Henderson (2021) found that the support structures in the APSBP are stronger than support structures in programs not affiliated with the APSBP. They provide examples such as strong orientation programs, and specific tutoring sessions for APSBP students [3]. Hodapp and Woodle (2017) indicate that APS helps their incoming students transition by creating programs that promote a sense of community and reducing the economic burden by providing financial support. APS also regularly monitors the progress of affiliated students, especially during their first year, and encourages bridging sites to establish regular staff mentoring[6]. Lastly, compared to the national student retention rate of 59%, the APSBP has a higher retention rate of 92% for their physics students[6].

B. Conceptual Framework: Community Cultural Wealth

Many studies in education research utilize frameworks that value cultural capital from the most privileged groups, which limits the discussion of the experiences of people from marginalized and minoritized groups. To intentionally oppose this practice, we utilize a conceptual framework that is guided by CCW[7]. CCW, which is rooted in Critical Race Theory, calls into question the white-middle-class standard and importantly highlights the valuable resources Communities of Color have to help them survive and adapt in existing institutions such as education. Furthermore, CCW works to prove that these communities do have the necessary skills to be successful, but institutions, such as education, do not appropriately value them. Thus, CCW is largely focused on shifting the idea from "fixing the students" to "fixing the institutions"[8].

Yosso breaks down CCW into 6 different categories that build upon Pierre Bourdieu’s idea of cultural capital. The six forms of cultural capital are: Aspirational Capital, Linguistic Capital, Navigational Capital, Social Capital, Familial Capital, and Resistant Capital. While each of the six categories has its particular definition, they all work hand in hand with one another, as seen in other studies that use this framework [7][8].

- **Aspirational Capital**: The ability to have hopes and dreams about one’s future, even in the face of barriers.
- **Linguistic Capital**: The skills attained through communication in different languages or styles. It relates most often to translating, storytelling skills, and communication through art and music.
- **Navigational Capital**: The skills and abilities an individual uses to maneuver through an institution. It is heavily influenced by social connections.
- **Social Capital**: The networks or groups an individual can establish. In education, these relationships are student-to-student or student-to-faculty support. These connections give students both emotional and navigational support.
- **Familial Capital**: The connections made through an individual’s family, community, or kinship. Examples in-
include direct family, sports teams, or religious groups.

- **Resistant Capital:** The knowledge acquired through oppositional behavior or behavior that challenges authority.

To be clear "culture refers to behaviors and values that are learned, shared, and exhibited by a group of people" [7]. Cultural capital as expressed by Bourdieu [9] and referenced by Yosso refers to "an accumulation of cultural knowledge, skills, and abilities possessed and inherited by privileged groups in society." The idea of cultural wealth acknowledges that there are multiple forms of capital that are learned and shared within marginalized communities that students use in educational settings.

**II. RESEARCH DESIGN**

The goal of this study was to determine which forms of CCW participants affiliated with the APSBP utilize in education and how the forms of capital used vary for different educational experiences. The goal of the interview protocol was to encourage participants to engage in storytelling about their experiences in the APSBP, including difficulties faced within their education and how they worked to overcome them. With these experiences, we could understand how participants in the APSBP utilized specific capital and how that changed.

**A. Data Collection**

Students who participated in the APSBP for at least one year at one of the original five APSBP sites were eligible to participate in this study. To find participants, we asked APS for contact information for students involved in the program. To date, eight semi-structured interviews have been conducted with participants that self-identified as Black, African American, Hispanic, Latino, Chicano, and Xicano Rican born in Texas. Additionally, these eight participants also self-identified as female, male/man, and non-binary. The interview protocol consisted of eight questions/prompts and follow up questions that were based on the interviewee responses.

1. Tell me how you became interested in physics.
2. Tell me about your decision to go to graduate school in physics.
3. Tell me about your decision to apply to the APS Bridge Program.
4. What kinds of challenges have you faced during your time in your graduate program?
5. What things have helped you overcome those challenges?
6. Have you made or helped to make any changes in your department?
7. How much control have you had in your graduate experience?
8. If you were to face a threat or barrier to your success in graduate school what are the things you would do to address that threat or barrier?

We wanted to encourage the interviewees to honestly share their experiences. Thus, we did not gear our questions toward any specific form of capital.

**B. Analysis**

Analysis of the interviews began with the creation of interview transcripts. Tempkin and Cochran then coded the first three interviews using a yes/no for whether or not a particular code was present. A total of 18 codes were possible based on the six forms of capital included in CCW and three different time frames: pre-college experience, when deciding to participate in the bridge program or graduate school, and during graduate school. The coders initially intended to have separate categories for the decision to go to graduate school and the decision to apply to the APSBP. However, after reading the transcripts and prior to coding, the coders recognized that most participants discussed making the decision to go to graduate school and the decision to apply to the APSBP simultaneously. Thus, these were combined to be one time frame. This resulted in the 18 possible codes seen in Table 1. After comparing a spreadsheet of codes for each of the interviews, Cochran and Tempkin discussed any differences in coding. In interview 1, there was 100% agreement for all 18 possible codes. In interview 2, there was 88.9% agreement, and in interview 3 there was 67% agreement. Based on this first discussion, Tempkin and Cochran re-coded the first four interviews. The overall inter-rater agreement post discussion was 100%.

Inter-rater agreement is one method for researchers to confirm that they are conducting analysis in the way they intend. We found that calculating inter-rater agreement - the degree to which our codes were identical given that the option was either yes/no for the 18 codes - was more beneficial than inter-rater reliability, which measures the extent of variability within a measure [10]. The method we chose to determine inter-rater agreement was a percentage based on the number of codes on which we agreed (as yes or no) for each transcript and the number of codes for which we could have agreed, 18 in all cases. The choice for determining inter-rater agreement is often a personal choice based on the research design [10]. In our case, it was helpful to determine inter-rater agreement as Cochran was initially more likely to only code an experience as illustrating one form of capital. Post discussion, Cochran was more careful to check for all forms of capital in the experiences shared during the interviews as sometimes a particular experience illustrated more than one form of capital. This change in Cochran’s method of coding resulted in the higher level of inter-rater agreement among the coders.
TABLE I. Tempkin Analysis Chart for Interview 1

<table>
<thead>
<tr>
<th>Capital</th>
<th>Pre-College</th>
<th>Decision Making</th>
<th>Graduate School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirational</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Familial</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Social</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Navigational</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Resistant</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Linguistic</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

III. RESULTS

Our study showed that participants within the APSBP utilize all six forms of capital during their educational experiences. Table 2 includes selected excerpts from the interview transcripts to illustrate the various forms of capital. One trend seen in the data was that the majority of pre-college codes were familial and aspirational capital. There were no pre-college experiences coded as resistant or linguistic capital. Secondly, most participants utilized navigational and social capital during the decision/application process of graduate school and/or the APSBP. Lastly, for all interviews familial, social, and navigational capital were illustrated in relation to experiences during graduate school.

A. Additional Findings: Familial Capital

When conducting qualitative research it is important to listen to the data. This requires noting what is there even if it does not directly address your research question. From listening to the data, we found that familial capital was a vital form of capital utilized by the participants throughout their experiences. Several pre-college experiences were coded as illustrating familial capital and all of the transcripts included experiences during graduate school where familial capital was utilized. The authors noted that during the interviews, there was a clear change in the physiological state of the participants as they discussed the role of family and the impact of graduate school on their connection with family. One participant, who used the pseudonym Amaru, said:

I’m the oldest of my siblings. My mom needs the help, and it was always hard to not be around. At the time I left for grad school, my little brother was just born. So he was a baby, and he grew up without me. He grew up without my accent, without my way of being, and so it always felt like I was missing out on the development of my family. ...

Also being away from my family, my great grandma passed away, and she was the last person of her generation. Like, that was my last great grandparent alive and at this point I was trying to figure out my roots and figure out where I come from because I had to learn about myself, and I never got a chance to ask her about her grandparents ... So like if this was the movie Coco, like that generation is now gone you know what I’m saying. So not having the opportunity to be around my family to ask those questions of where I come from, it’s a lot of information and a lot of knowledge that grad school would never, never ever provide for me.

TABLE II. Examples of Capital Coding

<table>
<thead>
<tr>
<th>Code</th>
<th>Interview Excerpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirational</td>
<td>&quot;I thought it was really interesting that we could use math and science to describe the physical world around us. After that I was like, alright I think I wanna do physics.&quot; (Interview 2)</td>
</tr>
<tr>
<td>Linguistic</td>
<td>&quot;So it was like I was missing a part of me and not being able to speak that language... the other language I speak it’s hip hop, so I was able to get some of that off and speak that language ... I was able then to make my own music which I then laced with a little bit of physics, because again, I do speak physics.&quot; (Interview 8)</td>
</tr>
<tr>
<td>Navigational</td>
<td>&quot;Ultimately, I decided to switch and do the research with the new advisor the summer before the qualifying exam.&quot; (Interview 1)</td>
</tr>
<tr>
<td>Social</td>
<td>&quot;Oh, it grew as I met people! ...there was a student orientation for all the incoming students, and the chair of the department was really helpful in making sure I had access to everything I needed and knew where to get things like groceries and...where to go to get a haircut. He’s always been a really good support.&quot; (Interview 3)</td>
</tr>
<tr>
<td>Familial</td>
<td>&quot;I probably called my parents an average of once a day. Also, my girlfriend. Video calling too...it helped ease any feelings of loneliness that I had.&quot; (Interview 3)</td>
</tr>
<tr>
<td>Resistant</td>
<td>&quot;I would do well in the class and would exceed the professor’s expectations for me in the class. I guess a motivator to do better in the class, now Imma show the professor I’m capable of doing well in your class.&quot; (Interview 4)</td>
</tr>
</tbody>
</table>
Multiple participants discussed the impact of being away from family. During Amaru's interview, Amaru mentioned that other students in the APSBP were also struggling with this issue and that they were trying to bring people together who faced a similar challenge. Amaru discussed efforts to "bring together other people who left their families behind, to pursue a graduate education."

B. Additional Findings: Socioeconomic and Ethnic Background

In our analysis, we noticed that socioeconomic background was a notable marker of difference [11]. We did not specifically ask participants about their socioeconomic background, but socioeconomic backgrounds among participants clearly varied. For example, one participant mentioned that her mother was a medical doctor, while another participant mentioned that their family members were migrant farm workers. So, we cautiously assume that the socioeconomic backgrounds of these two participants are different.

Differences in ethnic background and socioeconomic background caused some participants to be isolated - even within the bridge program. For example, one participant, who used the pseudonym Jorge, said

Here, it's like I'm not connected with people. ... it's very difficult for me to even interact with other people. ...I do not connect with other graduate students, including graduate students who are from other minority groups and are in the bridge program. I just don't have much in common [with them]. Yeah, I think that has been the most difficult. Not the classes, not the research, it's just interacting with graduate students including the ones from the bridge program.

For other participants, sharing a similar socioeconomic background helped them to connect even when they came from different ethnic backgrounds. One participant contrasts the connection between bridge program participants with other graduate students. He said,

All of us are either People of Color or Black, and often we're ...first people in our family to go to college and to have any type of degree. A lot of us ...we come from poor economic backgrounds, and maybe not all of our parents - thankfully mine do! But maybe not all of our parents support even us being in a grad program. Things like that, I would say. It's hard to find those commonalities between the other grad students as a whole.

IV. DISCUSSION

This ongoing project has yielded important preliminary findings. We found that all six forms of cultural capital were evident in the experiences shared by the study participants. CCW is a powerful way of acknowledging and understanding the various resources that students from minoritized ethnic/racial groups bring to their educational experiences.

Given that this is a qualitative study it is appropriate to explicitly state the trends that we found so that the reader can determine the transferability of results to other populations and contexts [12]. We found that pre-college experiences included a lot of aspirational and familial capital. This seemed reasonable as many students develop interest in science careers early in their educational journey due to family encouragement. We found participants utilized navigational and social capital most when deciding to go to graduate school. This may be because the decision-making process involves figuring out how and when to apply, and how graduate school might be financed [13]. Finally, participants used multiple forms of capital when faced with challenges in graduate school. These experiences presented unique challenges for participants, requiring them to incorporate all of their resources to overcome those challenges.

Additional findings included the importance of familial capital in overcoming challenges during graduate school and also the impact that the graduate school had on familial relationships. This finding led us to more questions. What are the implications of "leaving family behind to pursue graduate education?" What can be done to mitigate the negative impact of this process? What changes in policies and practices in education might eliminate this problem?

Prior to submission of this paper, Cochran presented these findings to bridge program students at one of the APSBP. They confirmed that these findings resonated with their experiences. Future work will include continuing the member checking process by sharing an expanded version of our findings with participants and soliciting feedback via written form and interviews.

ACKNOWLEDGMENTS

The authors would like to acknowledge the participants, Dr. Danielle Walker who assisted with data collection, the Inclusive Graduate Education Network, and Dr. Corey Ptak who assisted with the research design. This work is funded in part by NSF Award# 2050950 and #1834528. The findings included do not necessarily reflect the views of the NSF.

[1] R. Barthelemy, M. McCormick, and C. Henderson, Gender discrimination in physics and astronomy: Graduate student experiences of sexism and gender microaggressions, Physical Re-


Inequities and misaligned expectations in PhD students’ search for a research group

Mike Verostek (he/him)

Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627 and
School of Physics and Astronomy, Rochester Institute of Technology, Rochester, New York 14623

Casey W. Miller (he/him) and Benjamin M. Zwickl (he/him)

School of Physics and Astronomy, Rochester Institute of Technology, Rochester, New York 14623

Joining a research group is one of the most important events on a graduate student’s path to earning a PhD, but the ways students go about searching for a group remain largely unstudied. It is therefore crucial to investigate whether departments are equitably supporting students as they look for an advisor, especially as students today enter graduate school with more diverse backgrounds than ever before. To better understand the phenomenon of the group search process, we use a comparative case study approach to contrast important aspects of two physics PhD students’ experiences. Semi-structured interviews with the students chronicled their interactions with departments, faculty, and the graduate student community, and described the resources they found most and least helpful. Our results reveal significant disparities in students’ perceptions of how the group search process works, as well as inequities in resources that negatively influenced one student’s experience. We also uncover substantial variation regarding when in their academic careers the students began searching for a graduate advisor, indicating the importance of providing students with consistent advising throughout their undergraduate and graduate experiences.
I. INTRODUCTION AND BACKGROUND

Current data indicates that the retention rate of physics PhD students is approximately 60%, with attrition from PhD programs disproportionately affecting traditionally underrepresented students [1–3]. Leaving a PhD program can adversely affect students’ mental health and financial well-being, and early departures impact graduate programs that must invest time and resources into recruiting and supporting new students [4, 5]. As the number of physics PhDs granted across the US grows and the population of physics graduate students becomes more diverse than ever before [6], investigating the underlying factors behind high attrition is imperative.

Previous studies on graduate attrition across STEM and non-STEM disciplines indicate that a negative advising relationship is a key factor that motivates students to leave [4, 7–10]. At present, much research on graduate advising relationships has sought to identify which qualities are indicative of productive mentorships [11–14]. Despite the fact that finding an advisor is often cited as a crucial decision for PhD students to make [15, 16], existing studies generally do not examine the process by which these relationships form. PhD students often find navigating their first year of study to be difficult [17], and if the group search process is difficult for students to navigate, they may be less likely to find a group that provides them with a fulfilling research experience.

In one survey of approximately 4,000 STEM and non-STEM PhD students, students who reported being satisfied with their advisor tended to take more factors into account when looking for a research group than students who reported advisor dissatisfaction [18]. This suggests students’ access to information while searching for research groups is important in their eventual satisfaction in graduate school. However, it remains unclear how students go about gathering information about groups, what they prioritize, and which resources they find most helpful in that process. It is also unclear if there are differences across student demographics.

We investigate the phenomenon of the group search process in the specific context of physics PhD programs using a comparative case study approach. By contrasting two cases of physics PhD students searching for a graduate research group, we sought to answer the overarching research question of, “How do the experience and process of seeking and finding a research group vary among physics PhD students?” Within this overarching question, we also asked the following sub-questions: “What are some of the major questions and concerns that students have while looking for a research group?” and “What helps students answer these questions, and what gets in their way?”

II. METHOD

Case studies leverage specific cases to generate in-depth illustrations of an issue or problem [19]. Here, the cases for comparison are the narratives of two physics PhD students who recently completed their search for a graduate research advisor. Although case studies typically involve collection of data from multiple sources (e.g., artifacts, direct observation), our analysis is solely based on interviews with the two students. These were conducted as part of a larger study that includes over 40 interviews and is intended to characterize how students experience the process of finding a research group. However, due to the limited space available we opted to select two particular students with divergent perspectives on searching for an advisor and analyze them separately from the others. These two cases vividly demonstrate the impact that differing levels of access to resources can have on students’ graduate experience, thereby highlighting several important ways departments must better support students during their search for a research group.

Our semi-structured interview protocol was inspired by cognitive task analysis (CTA) methods [20, 21] and Dervin’s sense-making method [22, 23]. These methodologies are designed to elicit detailed descriptions of interviewees’ thoughts and actions as they recount how they progressed toward a goal, in this case joining a research group. Moreover, CTA and sense-making focus on specific life experiences that are bounded in time. These features make our data amenable to a case study analysis approach, since cases must have clear boundaries and great depth.

The protocol is broadly broken into three stages. First, we gathered a timeline of Steps that students took in their search for a research group. Students defined the start and end points of their stories, so these boundaries varied based on each student’s individual experience. Examples of common steps included “Applying to graduate school” or “Attending visiting weekend.” We then asked students to go through each step and identify their major Questions regarding their search for a research group at that point in time. This allowed us to gain insight into which aspects of the process drove confusion and uncertainty. Lastly, we asked students to identify any sources of Help that allowed them to resolve their question or concern, as well as any obstacles that Hurt their ability to move forward. This stage allowed us to understand the thoughts, actions, and events that helped or hurt students’ ability to navigate their search for a group. Once completed, interviews were transcribed and edited for grammar and clarity.

Next, focusing on only the two cases for this analysis, we coded the two students’ transcripts using a priori codes that aligned with the stages of the protocol: Steps, Questions, Helps, and Hurts. By associating Questions with particular Steps and Helps/Hurts with particular Questions, we generated timelines of each student’s individualized experience. From these timelines as well as repeated reading of the original transcripts, we wrote narrative summaries of the two students’ stories. These narratives helped us to begin identifying themes within each case, and eventually formed the basis for the case descriptions reported in the results. Lastly, we contrasted the Questions, Helps, and Hurts between the two students. This allowed us to make specific comparisons between each student’s sources of confusion, as well as the resources they used to navigate their advisor search. These cross-case
analyses allowed us to generate several general observations and takeaways, which we discuss in Section IV.

III. RESULTS

This section describes the cases of PhD students Alex and Brianna (both pseudonyms). Alex is a first year physics PhD student at a public doctoral university with Very High research activity. He self-identifies as Male and Hispanic. Alex attended an elite private undergraduate institution, and one of his parents holds a PhD in a STEM field. Brianna is a second year physics PhD student at a public doctoral university with Very High research activity. She self-identifies as Female and Black, and noted that she was the first in her family to attend graduate school. Her undergraduate degree was from a large public institution with Very High research activity. These two PhD students had drastically different experiences joining a research group in graduate school. For Alex, “things worked fantastically well,” as he matched with his advisor and began research the summer before graduate school started. On the other hand, Brianna indicated that “the process for me was pretty rough,” and her search for an advisor lingered into her second year of graduate school. Alex and Brianna’s individual narratives will provide a basis for identifying factors that led to their divergent experiences.

It is notable that COVID restrictions undoubtedly had some impact on both students’ their experiences. Brianna was a junior in Spring 2020, meaning that numerous REU opportunities were canceled during the summer of her junior year. Meanwhile, Alex was a sophomore in spring 2020, allowing him to do an REU during the summer of 2021 when many restrictions had been lifted. Still, neither student explicitly cited COVID as a major source of adversity, only noting that some restrictions exacerbated existing issues they experienced.

Case 1 (Alex): For Alex, the search for a graduate research group began the summer following his undergraduate junior year when he participated in an REU at his top-choice graduate school. Being in this position gave him access to graduate students and prospective PIs with whom he could conduct lab tours. “I basically just asked the grad student I was working with on the REU ‘Hey, do you know anyone in these two groups?’ And I looked them up and emailed them and said, ‘Hey, I’m a student, I will be applying to grad school next year, I like your labs. Do you have time to show me around?’” Following lab tours with graduate students, he recalled following up via email with the lab PIs, telling them “I am a student that will be applying that’s interested in your group. I did a tour with ‘name,’ and I really liked ‘project.’ Can we meet to talk further about whether or not you’ll be hiring students? My CV is attached.” Alex reported that during the summer, “I also met with the chair of the physics department and said ‘I really want to come here... Advice on applying effectively?’ And you know, once you get your foot in the door with the whole, like, I’m [a] physics REU student, the doors kind of open to you.” Alex said that his REU allowed him to “tailor” his application to his top school.

He also said that he “screened” potential graduate schools by research interests, looking to align his interest in experimental optics with groups at each school. He noted that “I got the advice about applying to places from my advisor saying, be persistent, make connections, network, do everything you can before the application goes in to make sure they know your name.” Alex was admitted to his top choice school, and expressed that at that time he felt “relaxed” because “I have a place to go. I like the people there. They want me to work for them.” During visiting weekend, Alex said that while most students were “deciding if they wanted to come to [the school],” he was spending his time narrowing down his advisor search. He again received advice from his undergraduate faculty advisor, this time regarding what to look for in prospective research groups: “I asked them, ‘Hey, what should I keep an eye out on for red flags?’” Moreover, the access he received as an REU student allowed him to more easily reach out to professors at the school a second time; in some cases, professors reached out to him. “That’s where I got to ask about work life balance, attrition rate, flexibility with choosing mentors. Are there rotations or are there not rotations? What does funding look like? How big are my classes? Can I take classes outside of physics? What do students do?” He noted that “It helped that I’m not the first person in my family to go to graduate school” when he was coming up with questions to ask PIs. He also asked again whether each PI was planning to hire students in the upcoming year, although Alex’s funding concerns were eliminated when was awarded a graduate fellowship.

Having the advantage of being able to meet professors during his senior year meant that Alex felt he had “so much time” to schedule meetings with graduate students, who he said he considered a vitally important part of his search process. In meetings with graduate students, Alex asked questions about lab culture, advisor support, work life balance, friends, classes, and general quality of life. “All of those questions, the culture and the like, were answered when I literally met with the grad students. I spent two and a half, three hours going through... I mean that’s why visiting is so important. But that’s also why having months instead of just a couple of days, where I could plan with so much time, if I have more questions I can come back in a few days.”

Alex was particularly interested in two research groups at his chosen school, and he described how “universally positive” feedback from both graduate students and the PIs made him “comfortable” with wherever he chose to go. Ultimately, he was “sold” on one of the projects when a PI told him, “If you’re also excited about doing this, which it sounds like you are, then let’s do it together... Like he basically said, ‘You are my first choice to work on this project.’” That advisor also went to Alex’s undergraduate institution for his postdoc, which contributed to their positive connection.

Reflecting on his group search process, Alex suggested that his experience might be “atypical” and “somewhat idealistic.” Overall, he said that he often felt as though departments “leave the student to fend for themselves. You figure it out.
We’ll help you if you ask, but we’re not really going to set up structures that promote your ability to efficiently meet with laboratories and students.” Despite this, Alex attributed his perceived success in the process to “find[ing] someone in a key position” to help “move the ball forward.”

**Case 2 (Brianna):** In selecting where to attend graduate school, Brianna highly valued her institution’s proximity to family and its superior financial package. She recalled, “I originally came to grad school with the idea that I’d be doing research in astrophysics. And I chose [this school] not because of the research, but because of the proximity... I saw that they did have a good like three or four astrophysicists here, so I thought I’d be able to do research in astrophysics.” However, upon arriving in her first year, Brianna reported difficulty navigating through graduate school: “Coming into grad school, I didn’t know anything about grad school. I’m the first in my family to go to grad school. So I had no idea what I was doing.”

In particular, Brianna felt that her physics department “didn’t really place an emphasis on finding a research advisor.” She remembered feeling unclear as to “whether or not I had to find [an advisor] right away, as opposed to like, waiting out and focusing on trying to pass the courses first. Because I feel like doing all of that was a little bit stressful.” Brianna was concerned that she might be falling behind, but also felt “too scared” to ask for guidance from faculty who “hold the fate of your career in their hands.” She remembered thinking that she “just didn’t know what was appropriate to ask and what not to ask. Like, I don’t know what I should have known coming in.” Meanwhile, Brianna said the pandemic had exacerbated her feelings of isolation from other graduate students, who she felt would have helped her “put things in perspective.” This prompted her to look online for advice, recalling how, “I would just go onto grad students subreddits and be like, okay, is this normal? Because it was during the pandemic, so I couldn’t really talk to anybody.” Although some online resources were helpful, she said they more frequently made her “[feel] bad” because students online “have everything set up and they’ve only been in grad school a month... like, I don’t know why I’m not as far along as everybody else.”

Shortly before the end of Brianna’s first year the department-level faculty advisor informed her that she needed to find a research group by the start of the summer, which was an unexpected development. “It was kind of blindsiding me because the general student advisor had initially said don’t worry about it, you have plenty of time to find a research advisor... I was under the impression that could wait until my second year.” At that moment, Brianna said she felt “scared” and “taken aback.” She set up a meeting with one of the astrophysics professors in the department, but he told her that he had taken on a new student just the day before, and was unable to take another.

In June of the first year summer she reached out to another astrophysics professor who informed Brianna that he might have funding to take on another student, but was not certain. Although this faculty member had neither accepted her nor supported her financially yet, she described feeling “bound” to his group. She explained, “I didn’t know if I was allowed to reach out to other people during the time waiting for him to come back to me.” In September, the professor told her that he did not have funding. “In the most plain terms he just kind of wasted my time for like, three or four months, when I could’ve used that time to talk to other professors. So I ended up talking to two more astrophysics professors asking if I could be in their group. And both of them said no, because they didn’t have funding for another grad student. So that’s all four in the department.” Having exhausted all of her options in the school’s astrophysics community, Brianna was uncertain about whether she should stay in the department.

At the time, Brianna felt her difficulty finding an advisor was due to a fault of her own, saying “Personally, I thought I was cursed. Like I slighted the department in some way, thinking that I had offended someone and I was being blacklisted from the research advisors. It didn’t make any sense. And I was just really having a hard time for a minute because I was like, I don’t know what I’m going to do now, if I can’t stay in the program.” By then, she was aware that many of her peers had found advisors, which added to her feelings of isolation, saying “I’m suffering alone. And I’m too embarrassed to admit it.” Brianna reflected that her journey in graduate school had been “just really depressing at points...It was like, I guess I’m not supposed to be here. Because everything is making it seem like it’s not working out.”

In what Brianna described as one last “desperate plea,” she approached her former E&M professor to ask if he would be able to advise her. She figured that she could “learn to like” some part of his research, but maintained worry that would not be the case. Still, she joined his group, motivated to persevere in her PhD. “He was like, ‘Yeah, we’ll be able to find something for you. You might be a TA a little bit longer, but we’ll figure something out for you.’ And he did his best to give me a project that was still kind of astrophysics related.” Brianna is currently still in his group and believes his willingness to adjust his own research agenda to better fit her interests has allowed them to work well together.

Brianna summed up her experience by saying “I was under the impression that I’d get a little bit more advising help, like how to navigate through grad school. And like I said, I really just had to figure it out on my own.” She elaborated, “I think I made the mistake of trying to place my trust into the professors, when I probably should have been trying to make connections with other grad students. Because they know what’s going on. And if you’re able to successfully do that, I feel like then they can actually tell you the truth.”

**IV. DISCUSSION**

Despite being largely absent from prior research on graduate education, these cases demonstrate the significant ramifications the group search process can have for new graduate students. This indicates that the search process itself may
be an important source of graduate attrition. While Alex de-
scribed feeling like advisors “want[ed] me to work for them;”
being rejected by several groups made Brianna feel “cursed,”
“alone,” and that “I’m not supposed to be here.” Rejection
from potential research groups, perceived by Brianna as a per-
sonal failing, clearly influenced Brianna’s sense of belonging
and drove her to consider leaving the program. Not only that,
but the numerous advantages Alex enjoyed while navigating
the search process allowed him to begin conducting research
a full year before Brianna, potentially shortening his overall
PhD timeline, and resulted in his ability to choose the exact
research lab he wanted to join.

Due to the highly impactful nature of the group search pro-
cess, it is critical for students of all backgrounds to be sup-
ported in their group search, and it is incumbent upon depart-
ments to provide that support. Indeed, we wish to emphasize
that the cases presented here are not meant to establish guide-
lines for what students should do while searching for a group,
but rather to highlight that departments must better support
students during their search for a research group.

Departments should have structures to clarify expecta-
tions and guide students in their group search: Both Alex
and Brianna reported that they did not feel their departments
provided guidance and support when it came to finding re-
search groups. However, Alex’s background as an REU par-

ticipant, undergraduate at an elite institution, and son of an
advanced degree holder gave him sufficient knowledge and
agency to navigate the unstructured system. Meanwhile, Bri-
anna felt like she “didn’t know anything about grad school”
and was constantly wishing for formal guidance about her
timeline, where to seek help, and norms of communication.
Departments that lack structures to accommodate students
like Brianna favor the success of students like Alex.

Recognizing that students enter graduate school with a di-
verse array of prior experiences, departments must imple-
ment programs to give all students the best chance of enter-
ing into a fulfilling advising relationship. Although detailed
recommendations are beyond the scope of this paper, the in-
equities revealed between Alex and Brianna’s cases begin to
provide insight into the ways departments must improve to
make this happen. For example, departments must strive to
provide clearly accessible sources of information that eluci-
date expectations for students. This might include informat-
on such as when in the PhD students are expected to join
a group and where funding is expected to come from, both
aspects of the process that Alex was able to anticipate but
Brianna expressed difficulty figuring out. Closely tracking
student progress throughout the degree is also essential.

Graduate coursework should be concurrently reformed
to more strongly emphasize finding a group: Research
comprises the majority of the graduate school experience,
which would seemingly make the process of joining a group a
top priority for first year students. Yet Brianna felt conflicted
as to whether to focus on coursework or find an advisor, and
cited coursework as actively hindering her ability to seek out a
research group because of the amount of time she dedicated to
classes. Although the purpose of graduate coursework is os-
tensibly to prepare students for research, Brianna’s case indi-
cates that the present emphasis on coursework in the first year
can actually hurt students’ ability to figure out what research
they want to do (this was never an issue for Alex, who felt he
had “so much time” to meet with groups and who found his
research group before classes began).

Classes are undoubtedly the most familiar and structured
aspect of the first year of graduate school, and are the only
place where students receive a grade for their performance.
Thus, for students feeling unsure of how to navigate graduate
school, the existing structure sends the message that focusing
on classes should be their top priority, perhaps at the expense
of other responsibilities like finding a group. It is therefore
important for departments to reform existing graduate curric-
ula to assure that coursework is not in tension with students’
ability to join a research group.

To this end, programs should consider ways of explicitly
integrating students’ advisor search into the graduate curricu-
"lum. Combining these features of the first year experience
would help alleviate students’ perception that they have to
sacrifice time looking for a group in order to focus on classes.
Moreover, revising course content to highlight available re-
search in the department and allow students to explore poten-
tial areas of interest would more closely align with the
goal of a curriculum that prepares students for their future
research. To assuage time constraints, courses could also be
spread more evenly throughout the PhD timeline to give stu-
dents more time for research and interest exploration during
their first two years.

Clarifying funding expectations and increasing trans-
parency of funding status are particularly important:
Alex and Brianna’s cases illustrate the critical importance of
illuminating the role of funding in students’ group search pro-
cess. Alex’s prior experiences allowed him to know that advi-
sors needed funding for him to join their group, whereas Bri-
anna expected to be able to teach in order to receive funding.
This made her search more difficult, as all four astrophysics
faculty informed her that they would not be able to support
her. Departments should therefore consider a means of mak-
ing advisors’ funding status more transparent to students.

Interestingly, like the first four professors she approached,
Brianna’s eventual advisor also initially indicated that he did
not have funding. The difference was he said he would “fig-
ure something out” for her, even if it meant teaching a bit
longer. It was unclear why the other professors could not do
something similar to accommodate her, and this aspect of Bri-
anna’s story points to a general ambiguity regarding advisors’
funding status that should be addressed; students must take
advisors at their word that they “don’t have funding,” which
gives advisors power to rebuff students with little explanation.

We thank the graduate students who participated in this
study. We hope their stories contribute to on-going improve-
ments to make PhD programs more humane, supportive, and
productive. This work is supported by NSF Award HRD-
1834516.


Computational modeling is by now a central pillar of modern science, yet it remains underrepresented in most introductory physics curricula at the post-secondary level. Recently, physics departments have begun incorporating computation into introductory physics courses, including those designed for life science majors. Such students form a significant proportion of enrollees in introductory physics courses, and will enter a technological world in which computational thinking is increasingly valuable. In this exploratory and qualitative study, we investigate how students connect to, and perceive the relevance of, computational activities built into an introductory physics for life sciences course. We use an ecological systems framework to understand and investigate relevance. We categorize the different ways in which students do and do not perceive the relevance of computational modeling in our physics curriculum. Our findings point to choices in course design that may increase relevance of computation for life science students, and indicate what mechanisms affect students’ perceptions of computational curriculum in these settings. In doing so, we begin to build a picture of how to incorporate computation into introductory physics courses.
I. INTRODUCTION

Calls to reform science education underscore the need to adapt undergraduate physics education to a changing world, focusing on present challenges and future opportunities [1]. This reform emphasizes transforming STEM learning and strengthening discipline-based education research by stressing the importance of a curriculum that not only promotes understanding of fundamental principles, but also equips students with practical skills applicable to their careers [2, 3].

The rising interest in blending computational modeling and Computational Thinking (CT) into science and engineering curricula is underscored by a growing body of scholarly research. Work by Chabay et al. [4] discusses the benefits and challenges of integrating computational material into an introductory, calculus-based physics course. Aiken et al. [5] investigate high school students’ understanding of CT and computational modeling on writing assignments and programming activities. More recently, Orban et al. [6] explored the meaning of CT in the setting of introductory physics, and Wellner et al. [7] develop a learning goal framework for CT in computationally integrated physics classrooms. These efforts highlight the need to better align STEM education with STEM practice, which already incorporates computation to a substantial degree, and to help students thrive in an increasingly technological world.

While these considerations are in many ways top-down, driven by instructors, administrators, and policy makers, it is important that students see the new computational content as relevant to their lives. This is a two-way street, and while students are often faulted for "not seeing the relevance," instructors should nurture environments which invite the student to consider connections to their lives [8]. Work by Walton and Sherman [9] indicates that personalizing learning activities according to "interests and experiences" improved performance in an Algebra 1 course, and a more recent work by McBride [10] explores customization in the context of physics for nonmajors. Lunke and Beichner [11] probed student attitudes towards computational modeling using spreadsheets in an introductory physics for life science (IPLS) course; however, there are limitations in the ability of attitudinal studies to probe relevance [8].

Our research addresses the following questions. What are the ways in which students find computational modeling relevant in an IPLS course? Further, is it possible to understand the sources and mechanisms of this perceived relevance in an Ecological Systems [12] (ES) framework? To answer these questions, we investigate student responses to computational modeling within the context of a studio-style course in introductory mechanics. We use an ES framework in our methodology and data analysis to probe, evaluate and identify themes around students’ perceptions of relevance. With insight from this research, we hope to inform curriculum design that makes computational modeling more relevant for students not only as a skill, but also as a tool for learning.

II. RESEARCH CONTEXT

The setting of our study is the first semester of a sequence of IPLS Studio [13] courses that follow the Modeling curriculum [14]. Students are typically in their second and third years of undergraduate study, and are overwhelmingly (~80%) concentrated in life science majors. In surveys given at the beginning of the course, more than half (~56%) of students report post-graduation plans involving medical, veterinary, dental, or optometry school, and another fifth (~20%) plan on graduate studies in life sciences.

The course is run in a "studio-style" where lecture and lab are combined into one interactive class session. In a typical semester, 48 students are enrolled per section, with 4 sections running concurrently. Within each section, students sit at tables in groups of 5-6. Collaboration is embedded throughout the course curriculum, including the computational modeling activities investigated in this study. Students can request frequent help from the instructional team, which includes several faculty and graduate teaching assistants, along with 8–10 undergraduate learning assistants.

Computational modeling using VPython [15, 16] in Glowscript [17], was incorporated into the curriculum through pre-recorded videos, five in-class group activities, homework problems, and several exam questions. No prior coding experience was assumed as most students reported never having seen code before. The in-class activities were dispersed throughout the semester, and though they were done in groups, each student worked on their own copy of code. Most of the activities involved modeling systems of both physical and biological significance. Students are given minimally working programs [7, 18] and asked to fix code to produce the desired results or tinker with the parameters. The first code, in week two of the course, asked students to move a ball with a constant velocity and then give that ball an acceleration. In week five of the course, students were asked to consider a model of ligand-enzyme interactions in terms of elastic and inelastic collisions. In week eight, students modeled paramecium motion and dissipative forces. In week eleven, students simulated bound states of molecules. Finally, in the thirteenth and final week of class, entropy and diffusion were discussed in the context of modeling ideal gases.

III. THEORETICAL FRAMEWORK

Relevance is a notoriously difficult concept to define in an educational context, especially in a way that does not diminish the student’s opinion and perspective. For an object, idea, etc. to be relevant to a student, it seems necessary that the entity links strongly with other aspects of the student’s life. This suggests a network representation of relevance, whereby an entity is relevant if students report many strong connections to other facets of their environment. These facets may include familial relationships, classes, clubs and organizations, career trajectories, and broad societal norms and expectations.

The sheer number of facets of a student’s environment necessitates an organizing principle. Thus, we augment our network-theoretic conception with an implementation of Bronfenbrenner’s ES framework [12], which was originally...
adapted for relevance in introductory physics by Sawtelle and Nair [8]. Figure 1 gives a schematic depiction of this framework for a hypothetical student. In the ES framework, a student’s environment is characterized by a series of nested layers, or "systems", with the individual at the center. Systems which are closer to the student are perceived as directly influencing, and being influenced by, the student. For operational clarity, we define the following systems in our study:

- The **Microsystem** comprises the most direct aspect of a student’s environment. This may include a student’s classes, friendships, and extracurricular activities.
- The **Mesosystem** is defined explicitly as connections within the Microsystem.
- The **Exosystem** is outside of a student’s direct experience but not an overarching societal level. For example, while a student’s MCAT preparation activities will fall in the Microsystem, the existence of the MCAT and other medical school admissions requirements are categorized into the Exosystem. Items in the Exosystem are relatively beyond the control of the student.
- For our study, the ** Macrosystem** is the largest possible layer and consists of entities at the large-scale societal or cultural level. For example, all of our students exist in a context with growing influence of big data and artificial intelligence. Broad economic trends towards information technologies may influence students’ choices and perceptions as well.

We do not consider the often-included "Chronosystem," or broad changes over time, because of the relatively short frame of time in which our study takes place.

IV. METHODS

Our study design is primarily qualitative, with the goal of uncovering themes from students’ experience. We view this as a necessary step for a more in-depth or quantitative study.

The data of our study is as follows. We created pre- and post-course surveys and journal reflections given in-class computational modeling activities to students in two sections of the course. Pre- and post-course surveys were obtained from 120 students and contained open- and closed-ended questions, including duplicate Likert scale prompts designed to gauge change over the course. Journal reflections, similar to the surveys, contained a mix of open- and closed-ended questions, and occurred during the course. We also conducted semi-structured interviews with three students near the end of the course. These interviews were approximately 30–40 minutes in duration and conducted over video calls.

We used the ES framework as a guide to probe relevance and its "source," e.g., the usefulness of computational modeling for a student’s career, major, physics, or everyday life. For instance, consider a prompt taken from a journal reflection activity following the moving ball activity of week two.

"Think of another situation where you might be able to build a computational model. This can be an everyday situation, or something related to another class. Describe that situation. Would computational modelling be useful, or would it be unnecessary in this situation – in what ways?"

Here, we are probing relevance of computational modeling to a student’s "everyday situation," which in our framework lies in the Microsystem. Similar prompts were given to probe relevance to other aspects of a student’s life, such as school or career, and the ES framework served to categorize them. We adopt a strategy employed in prior work [8] and avoid using the word "relevance" directly, since prior work has indicated that terms like "useful," "valuable," and "helpful" are more effective at probing relevance in written or verbal prompts. Similarly, we developed prompts using this approach for surveys and interviews.

After the end of the semester, all data were coded for emergent themes, performed independently by two of the authors (Watkins & Ivanov). Keeping the research questions and ES framework in mind, many of our codes take the form "computation’s relevance to ..." which signifies a connection a student makes between computation and a facet of their life. After a first round of coding, all authors met to discuss and compare the two sets of codes and categorization schemes. Not all codes and categories were considered relevant to the present study, and the authors collectively trimmed the list of categories from each code set. Of those remaining, there was general agreement between the two sets, and alignment with the research questions and theoretical framework. Using their respective sets of codes, the two authors performed a second, more thorough round of coding with emphasis on the most relevant categories to the study. QualCoder, an open source software, was used for the analysis [19].
V. FINDINGS

Our data provide snapshots of students’ attitudes and sense of relevance at different points in the course. We organize our findings according to major themes.

A. Initial Attitudes and Perceptions of Relevance

a. Students have inexperience and anxiety towards computation, yet some have determination to succeed.

Naturally, students came into the course with a variety of perspectives towards coding, and these are reflected in the pre-course survey. While 66% of students report no prior coding experience, 84% report using computers to plot, model, or analyze data (data analysis will be a recurring theme in students’ reflections on computers). When students were asked how much “anxiety do you feel about computer programming” in the course, 54% of students reported at least a moderate degree of anxiety (more precisely, these students gave a score of at least 60 out of 100, with 100 being most anxious). Notably, the distribution is bimodal, with 17.5% of students expressing almost no anxiety (defined as a score of less than 10 out of 100). One student wrote that “The most challenging part I believe will be the coding aspect when we get to that. I have zero exposure to coding so it will be completely new and is intimidating.” It was interesting to see a small population of students, who were anxious, also expressed determination. One student wrote “I sadly have never programmed before, but I’m hoping to catch on quickly and I’m excited to try.”

b. Computers are relevant for data analysis and visualization.

When students were asked “what role computers play” in “doing science in your major, area of interest, or future career,” most students reported some role rather than none, and their answers fell into natural clusters. The most salient clusters of responses were about data analysis, statistical modeling, and visualization. One author, in their parsing of student responses, recorded at least 31 distinct references to “data analysis” among study participants, without attempting a comprehensive count and classification, and was generally surprised by the consistency of these responses across different students. Indeed, this author had nearly four times as many coded texts for “data analysis” and “visualization” as for “simulations and modeling.” Connecting this to the CT practices outlined in [6], the students appear to be deriving relevance primarily from data analysis. Our observations align with the students’ stated background in programming (of those who have any). These students reported taking courses, previous or concurrently, by names like “Statistics for Scientists,” which are required courses for many life science majors and use Rstudio. They were already introduced to some form of coding and computational modeling, but specifically for data analysis, and have thus formed such connections in their Mesosystem.

c. Computers are relevant because they are everywhere.

Besides data applications, other students pointed to the general trend of increased technological penetration into our society (“Everything we do now involves computers,” one student writes). In the ES framework, we believe this to be in part due to the influence that the Macro- and Exosystems have on students attitudes towards technology and computers. For example, some students see computers as a valuable tool in the doctor’s office, through charting, medical scanning, or communicating. At this point, we had not asked about the role of coding specifically, but do in end of course surveys.

B. End of Course

a. Computational modeling is useful for learning how to code and to visualize but not useful for learning physics.

At the end of the course, not all students found computation relevant to physics, even if they found the activities enjoyable or useful in other respects — “I love coding and learned a lot but it didn’t help with my understanding in physics.” Although this student reported writing “simple” code prior to the course, and therefore already had an atypical amount of coding experience relative to our population, many other students who didn’t see a connection to physics managed to draw relevance — “This code was useful in understanding how coding works, but I did not find it useful for physics”. The interviews shed more light on why many students found learning how to code more relevant than using it to learn physics. One student mentioned that she was simultaneously taking a data analysis class and a computational modeling class, both requirements for her biochemistry major. She felt that the kinds of coding done in each class were “significantly different,” yet she said it was nice to see all the different languages and styles. Despite these differences, “I definitely think taking them simultaneously helped kind of learn the codes and learn how to use the programs at the same time.” While she found examples towards the end of class, involving thermodynamics concepts such as enthalpy and entropy, helpful and interesting, she ultimately “didn’t really see the relevance” of the coding to her physics learning. However, without the basics she learned here, she said “I totally would be I’d be lost [sic.] in the rest of my [computational] modeling classes.”

Thus, in the ES framework, it appears that the link between physics and coding is weak for many students, while for a few students, coding remains strongly connected via other nodes such as their major. Further, we observed that students who
saw visualization as the primary use of computational modeling activities found it less relevant for learning physics: "For me [coding in the collision activity] was not very useful, I had a very similar idea when looking at the drawn model, but I can see how it could help someone visualize."

**b. Coding is relevant for building problem solving skills.**
A few students found value in developing critical thinking and problem solving skills. "I do know I won’t be writing any code, but a big component of coding is finding the problem and fixing it. This is necessary in any career." Another student remarked that this skill was of more value to their future career than actually learning physics: "...what’s probably more helpful than the physics content is the ways that physics and coding teach me to think about our program... maybe coding will be helpful, but definitely thinking skills from the class will be helpful." We briefly note parallel responses in our data about learning physics: whether or not the content or techniques apply, it is useful for problem solving and critical thinking.

**c. Coding is useful as tool to calculate.**
Some students highlighted the ability to perform complex calculations with a computational model. One of our interviewees, a human biology major, had coding experience through a required statistics course, which officially used Rstudio, but in practice she could get away with using a TI-84 calculator. Prior to the course, she saw coding as an activity most useful for "computer-based" activities like making websites, but was starting to see value in performing complex calculations: "Like, recently we had an [drag and viscosity] equation, and it’s really difficult to do unless you know advanced calculus, but like we just put it in our code and then it gives us all the values out, which is super, super nice."

**d. Computational modeling as a tool to understand physics.**
Being able to translate between equations and computer code, tweaking the physics model, and in the process gaining better conceptual clarity is a very important reason to use computational modeling in the curriculum. Some students found the code useful in helping understand their physics model, specifically the effect of changing different variables: "This code was useful with experimenting how different factors like mass and velocity impact collisions." Another student remarked "This code was a deeper dive into modeling slightly more complex scenarios... By using a program to model this, I was able to see the calculations behind the actual motion, and see how altering those calculations affects the actual simulation and motion of the objects."

VI. DISCUSSION & OUTLOOK

In this exploratory study, we investigated the different ways in which students derive relevance towards computational modeling in an IAPS course. We find that students, broadly speaking, came to the course perceiving relevance in terms of data and visualization, and also pointed to current societal impacts of technology. Within our ES framework, this can be understood, since students from life science disciplines are likely to interact with computation strongly through a data-analytic perspective. We further see evidence of the Macro System increasing students’ sense of relevance, with "computers are everywhere" being a common sentiment in student responses. On the other hand, students did not always see the relevance to learning physics. Most students who saw the code primarily as a visualization tool did not find the activities as valuable compared to other ways of learning, such as physical experiments. Instead, some felt the code improved their problem solving skills, and others had motivations to learn how to code for research or other courses. In contrast, some students did find computational modeling useful as a tool to perform complex calculations and also as a tool to learn physics by exploring relationships between the variables in their physics models. In this course, students and instructors referenced computational modeling as "coding activities," and for future studies it would be interesting to see if using the phrase "computational modeling" would further encourage students to find relevance to learn physics.

Future studies could more explicitly probe the correlations between sources of utility (e.g., visualization, calculation, or variable relationships) and students’ perceptions of usefulness. Although the visual aspect of the coding activities may have helped novice coders get their footing, it may have unfortunately acted as a red herring, distracting students from the more significant uses of computation in physics, such as performing otherwise-intractable mathematical calculation and helping understand physics models. Future design of computational activities can benefit by more intentionally encouraging students to use computation as a tool for learning physics.

Our study involves primarily life science majors, who tend to be a large cohort in undergraduate physics classes. Engineering majors form another large population, and might be expected to have more coding experience or find more relevance towards computational modeling. Performing an analogous study to investigate relevance in this group would provide a more complete picture for a typical introductory physics course.

The heavy emphasis many students placed on data analysis and plotting in initial surveys suggests pedagogical strategies for enhancing relevance. Instructors looking to draw students into computational activities might lean on this background in the beginning and introduce other aspects of CT over time. The use of Jupyter Notebooks, rather than VPython with GLOWScript, might be better suited for this purpose. However, instructors should not assume students have programming experience, and need to find ways to incorporate computational lessons without increasing students’ cognitive load.

ACKNOWLEDGMENTS

We thank Vashti Sawtelle for the helpful discussion on the theoretical aspects of our research design, and Bryan Stanley for helping us prepare effective interviews. We thank the participants and mentors in the SUTL fellowship at Lyman Briggs College, with a special thanks to program director Rachel Barnard. This work was performed while Watkins was a NSF Graduate Fellow, and while Watkins and Ivanov were SUTL fellows.
ChatGPT reflects student misconceptions in physics

Sav Wheeler and Rachel Scherr
School of STEM, University of Washington Bothell, 17927 113th AVE NE, Bothell, WA, 98011

With many students turning to machine learning models such as ChatGPT for homework help, diagnosing the accuracy of the information these technologies purport across different fields is more important than ever. In this study, we find that ChatGPT’s responses to standard physics questions are consistent with an accurate theoretical understanding, but are often incorrect when applied to a given system, reflecting many of the same misconceptions as students.
I. INTRODUCTION

Released in November of 2022, ChatGPT is a trained AI (artificial intelligence) model based on the large language model GPT-3.5, with a focus on training for a dialogue-based conversational format [1]. The AI gained widespread popularity soon after its release, including among students, who frequently used the AI for academic help [2, 3]. The potential for student use of GPT as a writing tool has sparked fears among educators of a new wave of plagiarism and cheating [3], though text-checker tools have since been developed to mitigate this [4, 5]. ChatGPT has gained academic notoriety for passing or excelling at several standardized tests such as the SAT, GRE, and multiple AP tests [6]. However, these benchmark tests may not be representative of field-specific student understandings, and in several cases, ChatGPT has not only failed an exam, but performed worse on average than student responses [7].

Haensch et al.’s [8] evaluation of social media content relating to ChatGPT, in an effort to gauge student views on use of the program, suggests that students generally view ChatGPT positively in its applications and use for writing essays, generating code, and answering questions. Alongside this, their study suggests an under-representation on social media of ChatGPT’s capacity to generate misinformation in these applications. Bridging the gap between the content-critical academic perspective and opportunistic student perspectives on this technology will be essential, going forward, in ensuring students know the limitations of the tools at their disposal, and develop a degree of awareness of how generative AI technology can affect their learning for better or worse.

In this study, we analyze the performance of ChatGPT on the Force Concept Inventory (FCI) in order to gauge how this AI might affect student learning of key physics concepts by reflecting common-sense misconceptions. While this inventory is not necessarily representative of the complete range of topics and concepts in introductory physics courses, it contains a breadth of essential foundational concepts such as Newtonian laws, kinetics, and momentum. We focus on the categorical performance of ChatGPT on sets of questions in contrast to overall performance as well as a qualitative analysis of specific prompt responses that may be consistent with common student misconceptions.

The purpose of this study is to quantify ChatGPT’s potential to inform or mislead students who use the AI. The quantitative analysis can provide a general overview of the accuracy of GPT’s responses, while the qualitative analysis can provide a more in-depth look at what particular problem-solving skills or frameworks ChatGPT’s responses appear to employ, and whether these resources are consistent with students’ own.

II. METHOD

We prompted ChatGPT with questions based on the FCI. Many of the questions were based on diagrams, so written transcriptions of the diagrams were given for any question which depended on one, including any questions with answers that were a part of the diagram or other representative visuals. All prompts are included with ChatGPT’s responses on the website listed in the Data section of this paper.

Five standard trials were run for the exam, where all questions for a trial were given in the same chatroom, including question numbers and lettered options. Three “variant” trials were also conducted, which tested different experimental changes to the testing method. Variant trial 1 had ChatGPT answer the questions individually in separate chatrooms. Variant trials 2 and 3 included a preliminary “role-play” prompt to see if ChatGPT performs better or worse when taking on a given role, which were respectively:

- Variant 2: “Pretend you are a first-year physics student taking an exam. Answer the following questions with your letter selection. I’ll start:”
- Variant 3: “Pretend you are a physics professor creating an answer key for an exam. Write the correct letter answers to the following questions. I’ll start:”

III. RESULTS

Table I shows ChatGPT’s letter answers to each FCI question prompt, with incorrect responses labeled in bold. Cells labeled “N” indicate responses where no letter answer was given. Exact prompts used and ChatGPT’s written responses to each prompt can be found on the following website: https://lucinder.github.io/CapstoneStudy/StudyData.html

A. Overall Performance

ChatGPT scored generally poorly on the Force Concept Inventory ($\bar{x}= 51\%$, $\sigma = 8.5\%$), with a distribution of predicted scores falling within 44%-59% with 99% confidence. An FCI score of 60% is considered to be an entry threshold to Newtonian physics [9]. Notably, GPT had a slightly lower median score ($\bar{x} = 48\%$) than mean score, indicating a slight right skew in scores. Scores on variant trials were 10.3% worse on average than scores on standard trials.

B. FCI Cluster Performance

The original publication of the FCI outlines six “clusters” of concepts, as well as indicating which questions contain answers that rely on an understanding of each concept [10]. Based on ChatGPT’s responses to questions containing answers corresponding to one or more of these concept clusters, we recorded the metrics of ChatGPT’s performance on different clusters.

Table II shows ChatGPT’s average scores on each FCI cluster along with their number of questions ($n_q$), standard deviations, and 99% confidence intervals of scores. Fig. 1
TABLE I. ChatGPT’s letter responses to each FCI question. Incorrect responses are indicated in bold.

<table>
<thead>
<tr>
<th>Trial</th>
<th>Question No.</th>
<th>Score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C E A B D E B A A C B C B B E D D D E B B A C C</td>
<td>58.6</td>
</tr>
<tr>
<td>2</td>
<td>C E A A C D C A B B B B D A B B E A A A C B</td>
<td>48.2</td>
</tr>
<tr>
<td>4</td>
<td>C E A B D B E A B B A A B B B A A B N E A A D E A A A C C</td>
<td>44.8</td>
</tr>
<tr>
<td>5</td>
<td>C E A D D C D B A D B C B A A C B B C B D D B D A E A D C</td>
<td>44.8</td>
</tr>
<tr>
<td>Variant 1</td>
<td>C A D D C D B A D B C B A A C B B C B D D B D A E A D C</td>
<td>44.8</td>
</tr>
<tr>
<td>Variant 2</td>
<td>C E A A C C A A D B E B A A B C B B N E E A B C A C C C B</td>
<td>44.8</td>
</tr>
<tr>
<td>Variant 3</td>
<td>C E A A C D B A D B E B A A B N B B N A C A B C A A A C B</td>
<td>44.8</td>
</tr>
</tbody>
</table>

TABLE II. Key figures from ChatGPT’s cluster performance

<table>
<thead>
<tr>
<th>Concept Cluster</th>
<th>n_q</th>
<th>q̄</th>
<th>σ</th>
<th>99% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinematics</td>
<td>7</td>
<td>23%</td>
<td>15%</td>
<td>[9.0%, 37%]</td>
</tr>
<tr>
<td>Newton’s First Law</td>
<td>8</td>
<td>63%</td>
<td>13%</td>
<td>[50%, 75%]</td>
</tr>
<tr>
<td>Newton’s Second Law</td>
<td>4</td>
<td>19%</td>
<td>29%</td>
<td>[0.0%, 45%]</td>
</tr>
<tr>
<td>Newton’s Third Law</td>
<td>4</td>
<td>88%</td>
<td>19%</td>
<td>[70%, 100%]</td>
</tr>
<tr>
<td>Superposition Principle</td>
<td>4</td>
<td>63%</td>
<td>13%</td>
<td>[50%, 75%]</td>
</tr>
<tr>
<td>Types of Force</td>
<td>12</td>
<td>50%</td>
<td>10%</td>
<td>[41%, 59%]</td>
</tr>
</tbody>
</table>

FIG. 1. ChatGPT’s average score in each concept cluster.

also shows the average score on questions in each concept cluster by ChatGPT, with margins of error indicated by thin lines on bars. ChatGPT notably performed poorest on questions that probed concepts of Kinematics and Newton’s Second Law.

IV. ANALYSIS

ChatGPT’s consistent incorrect responses may result in students receiving incorrect or confusing information when using ChatGPT as a learning tool for physics. The following sections discuss the potential causes and implications of ChatGPT’s incorrect responses, and how educators can adapt to widespread student use of ChatGPT with these limitations in mind.

In some cases, ChatGPT’s responses to FCI questions represent accurate theoretical understanding of concepts in physics, and its explanations may be helpful to students. For example, when asked to compare the collision forces between a large truck and small compact car, ChatGPT explains,

"According to Newton’s Third Law of Motion, for every action, there is an equal and opposite reaction. Therefore, during the collision between the truck and the car, both vehicles will exert equal and opposite forces on each other. The truck exerts the same amount of force on the car as the car exerts on the truck."

In other cases, though, ChatGPT responses are likely to mislead learners. ChatGPT’s responses to kinematics questions reflected common student difficulties distinguishing velocity and acceleration, as well as a tendency to add velocities as scalars rather than vectors [10]. For instance, when asked about a constant-velocity hockey puck that receives a horizontal “kick” in the direction horizontally perpendicular to its existing path in a question probing students’ ability to interpret motion as vectors, ChatGPT explained,

"Since the surface is frictionless, the horizontal component of the velocity of the puck should remain constant throughout its motion. Therefore, the speed of the puck just after it receives the kick should be the arithmetic sum of the speed it had before the kick and the speed it acquires from the kick. Thus, the answer is the arithmetic sum of speeds v0 and v."

In another response to the hockey puck scenario, ChatGPT interpreted the instantaneous “kick” to be a long-lasting force that produces circular motion, which may reflect the
student common-sense misconception of confusing motion with force [11]:

When the hockey puck receives an instantaneous horizontal “kick” in the direction horizontally perpendicular to its existing path at point B, it will start to move in a curved path. Since the surface is frictionless, there is no force to slow down or stop the puck’s motion. The puck will continue moving with a constant speed but changing direction due to the force acting on it.

Some common student misconceptions concern object behavior in outer space, in particular the common sense misconception that objects are weightless in space or that Newton’s Second Law does not apply in the context of outer space [12, 13]. In response to a question about a rocket in outer space firing its engine with constant thrust at a right angle to the rocket’s initial motion, ChatGPT incorrectly states that the rocket’s speed is constant during the thrust:

As the rocket’s engine produces a constant thrust at a right angle to the line [of initial motion], the rocket will move in a circular path... Therefore, the rocket’s speed will be constant, as it moves with a constant angular velocity.

Overall, ChatGPT is an unreliable source of physics reasoning. Sometimes its interpretations veer toward the surreal: in a question regarding a ball on a string that a person swings towards the person. After an instantaneous horizontal “kick” in the direction horizontally perpendicular to its existing path at point B, it will start to move in a curved path. Since the surface is frictionless, there is no force to slow down or stop the puck’s motion. The puck will continue moving with a constant speed but changing direction due to the force acting on it.

According to Newton’s First Law of Motion, an object in motion will continue in a straight line with a constant speed unless acted upon by a net external force. In this case, the heavy ball is moving in a circular path due to the tension force in the string, which provides the necessary centripetal force. When the string suddenly breaks, there is no longer a centripetal force acting on the ball, and it will continue to move tangentially to its path at the moment of release. However, since the person is also moving in the circular path, the ball will also have a component of motion towards the person. Therefore, the path the ball will travel relative to the person is mostly tangential to its path at the moment of release, but curved inwards along the horizontal plane towards the person.

The AI may be interpreting the language of this prompt as similar enough to another “archetype” of question in this case, misinterpreting a question about horizontal centripetal force as a question about orbits-to attempt to answer the prompt as if it were responding to the other archetype. Though not necessarily reflecting a misconception, this may confuse students in regards to how the question should be interpreted.

The content of ChatGPT’s explanations in its prompt responses reflects many of the common-sense misconceptions of students, particularly those regarding kinematics, and may mislead students who use the AI as a study tool. Patterns of treating vectors as scalars, confusing momentum with force, and misinterpreting forces in the context of outer space are some of the main misconceptions which ChatGPT may reinforce when responding to physics questions.

V. STUDY LIMITATIONS

Many questions from the FCI relied on, or had answers that relied on, a visual representation or diagram. The public model used by ChatGPT, GPT-3.5, cannot process images, and as such, all diagrams and representations were translated into text to the best of our ability. A lack of access to diagrams may have affected ChatGPT’s scores on these questions, but until such diagrams are tested with GPT-4, it is uncertain how exactly the absence of diagrams would have affected ChatGPT.

All prompts were given to, and responses collected from, the ChatGPT March 23 release. Incremental updates in the AI may slightly change how it responds to certain prompts.

VI. RECOMMENDATIONS FOR EDUCATORS

ChatGPT offers accurate, conversational explanations of physics theory in some cases, and incorrect or misleading explanations in others. Students who use it in a physics context may benefit from its explanations, but also risk having common misconceptions reinforced. One potential educational application of ChatGPT is producing practice problems that require students to identify and correct a problem ChatGPT has solved incorrectly. ChatGPT’s responses are worded convincingly and often purport accurate theoretical explanations to supplement incorrect practical explanations, and such questions would therefore challenge students to think critically about the information presented in ChatGPT’s response and formulate a logical flow of operations for solving the same problem. This would allow students to further familiarize themselves with applying concepts in physics, as well as giving them practice in fact-checking in these fields.

In general, an outright ban of ChatGPT in the classroom is unlikely to dissuade students from using the AI for their own purposes outside of the classroom; thus, giving students the tools to reconcile their own understandings with information purported by the AI would be highly beneficial to students who do use the AI regularly, especially as this technology becomes more integrated in our daily lives in the future.
Examining faculty choices while implementing the Next Gen PET curriculum through Revealed Causal Mapping

Julia Willison (she/her)
Department of Physics, University of Central Florida, 4111 Libra Drive, Orlando, FL, 32816

Erin M. Scanlon (she/her)
Department of Physics, University of Connecticut - Avery Point, 1084 Shennecossett Rd., Groton, CT 06340

Jacquelyn J. Chini (she/her)
Department of Physics, University of Central Florida, 4111 Libra Drive, Orlando, FL, 32816

Next Generation Physical Science and Everyday Thinking (Next Gen PET) is a research-based, active-learning physical science curriculum for general education physics courses, with a focus on pre-service and in-service elementary school teachers. During pre-COVID introduction of this curricula at higher-education institutions across the country, instructors implementing the curriculum were recruited to participate in faculty online learning communities (FOLCs). In this project, we conduct a secondary analysis on transcriptions of recorded FOLC meetings through Revealed Causal Mapping (RCM), a qualitative technique to create causal maps composed of interconnected causal statements. This method allows us to examine instructors’ decision-making while implementing the Next Gen PET curriculum. In this paper, we share insights about how instructors decide to form groups, specifically group composition and frequency of group changes. We found that instructors may have their own preconceived ideas about the best ways to form groups, sometimes contrary to what current research suggests.
I. INTRODUCTION

Implementing new curricula takes time and effort from faculty. New curricula may try to be one-size fits all, but different faculty at different institutions have different wants and needs to best serve their students. Previous research has shown that when incorporating new curricula, instructors tend to make modifications to the curricular materials that are being implemented [1]. The Next Generation Physical Science and Everyday Thinking (Next Gen PET) curriculum is a university general education science curriculum with a focus on pre-service teachers [2, 3]. It was designed to be flexible to help accommodate variations in faculty and student needs [2]. Part of the flexibility of the Next Gen PET curriculum is the instructor’s ability to make explicit choices in their implementation of the curriculum [3]. This includes choosing a class style--lecture or studio--and which of the modules and activities to use [3].

To help faculty implement the Next Gen PET curriculum, faculty online learning communities (FOLCs) were created. Two to three faculty with experience working with the curriculum were assigned as facilitators to each FOLC. Each FOLC consisted of approximately 8-13 instructors.

The Next Gen PET curriculum developers and FOLCs emphasized the importance of several active learning strategies, including whole class discussion and group work. In this paper, we focus our analysis on discussions of group work in one FOLC, and specifically on the mechanisms behind creating groups of students and shuffling groups. All the participants in the FOLC we analyzed were using the studio-version of the Next Gen PET curriculum. Our research question was: What reasoning supports instructors’ decisions to implement a variety of groupwork strategies in the Nex Gen PET curriculum?

To answer this question, we utilized Revealed Causal Mapping (RCM), a mixed-methods analysis used to create causal maps composed of interconnected causal statements which can show paths of reasoning [4]. This method is particularly useful when looking at the logic of expert decisionmakers [4]. We consider instructors to be expert decisionmakers within the domain of their instructional experiences, and we look to elevate their experiences to inform curriculum and professional development.

A feature of RCM, like other mixed methods techniques is to “facilitate the transformation from qualitative inquiry to quantitative inquiry,” [5]. Teddlie and Tashakkori’s terminology for mixed-methods would likely classify RCM as a monostrand conversion design [6].

II. METHODS AND DATA

A. Data

The analysis presented here is a secondary analysis conducted on a pre-existing dataset consisting of verbatim transcripts of one Next Gen PET FOLC’s meetings during the 2017 – 2018 academic year, the first year these FOLCs ran. This FOLC met about once per week for one hour per meeting, with faculty participating every other week and facilitators swapping about every other week, so that at least one facilitator was present at each meeting. Author J.J.C. was a facilitator in a separate FOLC for this project. She also has used the Next Gen PET curriculum in her general education physics course for 8 years.

The focal FOLC was made up of thirteen faculty, including 3 facilitators. The group varied in gender (7 men, 6 women), race/ethnicity (9 white, 4 non-white), and institution type (8 Master’s degree-granting, 5 Ph.D. granting). Eleven instructors are represented in the data presented in this paper. The 2 instructors not represented here did not have any relevant causal statements within the dataset. The FOLC met 33 times in total. We analyzed transcripts for 22 of the 33 meetings, with analyzed transcripts sampled by convenience. The analyzed meeting transcripts span the entire school year. Previous research suggests that attitudes and content of FOLC meetings change across the course of the FOLC, so it was imperative that we encompassed the entire year with our analyzed data [7].

B. Methods

Our analysis was primarily conducted using the NVivo qualitative analysis software. The transcripts were divided among the three of us so that at least two coders analyzed each transcript.

Nine transcripts were utilized to develop and test our coding method in a two-stage process. First, four of the nine transcripts were used to develop the keywords to search for sections related to the focal practice of group work. This was done by reading through the transcripts for relevant sections and noting common words describing the instructional practice. The keyword list we developed is as follows, (* denotes any form of the word was used as a keyword):

Group, dysfunc*, norm, expect*, eval*, account*, partner, disc*, turn, present*, exam, project, focus, incentive, slow, fast, check, goal, collab*, dynam*, split, facil*, activit*, work, concentrat*, task, behind, share, commun*, behav*, listen, learn, together, public, everybody, engag*.

After the keyword list was developed and relevant areas of the four transcripts were identified, all three coders searched for causal linking words and identified the causes and effects together. This process allowed us to develop norms for identifying causal statements and their causes and effects. The next five transcripts were coded individually by two coders each to refine the norms and the coding process.

For the final thirteen transcripts, two coders individually used the keywords to distill the data down to the sections that discussed group work. Relevant sections were found in 20 out of 22 transcripts analyzed.
I. Cause and effect identification

The next step was to find causal statements within the relevant sections of data. Causal statements consist of a cause, a linking word, and an effect [5]. First, we searched for the causal linking words outlined by previous literature: if, then, because/cause, so, since, think/thought, know/knew, use, believe, feel/felt, and which [5,8]. We only searched for the linking words within relevant sections of each transcript. Then, once all the causal linking words in the relevant sections were identified, they were manually reviewed to determine the cause and effect linked by each word, if any.

There are some uses of the causal linking words that are not causal in nature (e.g., “I give the Lawson Reasoning Test and then I kind of try to put a group of four...” uses a temporal, not a causal, “then”). These statements were not included in our data set. Causal statements that were not relevant to group work were also not included in our dataset.

To build trustworthiness of our analysis, two individual coders coded each transcript. To ensure inter-rater agreement (IRA) after individual coding, the two coders met and compared each relevant causal statement found within the text. In the event of a discrepancy, the differing statement was discussed between the coders until an agreement was reached to keep or not keep the statement. If an agreement was not reached, the statement was brought to the third researcher, who had not coded that section, to finalize a decision on the statement. In total, 37 statements were discarded through IRA, with 7 of those being from the Creating Groups set. The Creating Groups set contained 71 statements out of the 216 statements kept after the IRA.

The statements were then sorted by the classroom practice being discussed, and the statements in each section were numbered. From there, the causes and effects of each statement were aggregated into larger grain size categories to be used as nodes in the maps. Two coders each individually aggregated the larger grain categories, one via Grounded Theory, by line-by-line coding, then aggregating those codes into larger grain size categories and the other via content analysis [9,10]. Then, the categories were discussed between the two coders and the third researcher. The aggregated categories had many similarities between the two coders, so the smaller set of categories was chosen, with a few changes being made to reflect ideas from the larger set. For example, “When I first mix up the groups, I look for all the weak students and I put them together in a group,” and “I fall into the group of people that leans towards homogenous groups I think,” both fall into the node Group same level students. The aggregated statements formed 43 nodes in total, some of which are just causes, some just effects, and some that are both.

2. Causal mapping

After the causal statements were fully identified and classified into wider causes and effects, each coder proceeded to make an initial causal map. The maps consist of a cause, effect, or combination cause and effect as a node; and lines connecting individual nodes with arrows pointing in the direction from cause to effect. Some lines may have arrows in both directions, denoting that the causal relationship goes both ways in the data set. The maps were compared for accuracy and agreement. From there, two researchers each created a second draft of the map, with statement numbers and cause and effect numbers at each node. The maps were compared once again for accuracy and agreement.

III. FINDINGS AND DISCUSSION

The set of statements coded under the “Group Work – Creating Groups” instructional practice ultimately formed two connected but thematically distinct maps, Creating Groups and Shuffling Groups. Eleven of the thirteen FOLC members are represented in these two maps (i.e., the FOLC members’ causal statements, derived from the transcripts, are included in the nodes in the causal map). The Creating Groups map focuses on the strategies discussed for making groups, while the Shuffling Groups map focuses on the frequency of forming new groups throughout the semester.

Some nodes were not connected to the larger maps. Most are single cause-and-effect statements; most are presented in Fig. 3. They will be referred to as “isolated RCMs.”

A. Creating groups

Creating Groups is centered around three large nodes, the instructional practices C. Evenly mixed group level and O. Group same level students, and the faculty reasoning H. Group works well. The next largest nodes are L. Randomized groups, F. Student dissatisfaction, and P. Improve student effort. Five of the isolated RCMs were explicitly related to the Creating Groups RCM (see Fig. 3).

It is unsurprising that the Creating Groups RCM is dominated by C. Evenly mixed group level, O. Group same level students, L. Randomized groups, and H. Group works well. Ultimately, the goal for instructors is to have groups in their class work well. As one instructor stated, “If you group them up … really seems to make all of them better, which makes me happy.”

This data shows that the path to groups working well is varied. Both C. Evenly mixed group level and O. Group same level students led to H. Group works well. Not only that, but they both lead to P. Improve student effort. Other curricula, such as Student-Centered Activities for Large Enrollment Undergraduate Programs (SCALE-UP), call explicitly for evenly mixed groups, with the reasoning that the top students in a group can help assist the lower students, leading to an overall improvement [11]. Additionally, research in physics education suggests that varying student ability in groups allows students to support each others’ learning [12,13].
FIG 1 Creating Groups causal map. Nodes in this figure are lettered A – S., left to right, top to bottom. The numbers in the parentheses denote (# of statements, # of faculty). Yellow square nodes are instructional practices, and indigo oval nodes are faculty reasoning. The color saturation of the nodes denotes its frequency in the dataset, with lighter colors having fewer statements.

It is interesting to note that L. Randomized groups is not connected to H. Group works well in this analysis but was connected to F. Student dissatisfaction by a two-sided arrow. Looking back at the causal statements used to make up these nodes, we find that the instructor understood that the students were dissatisfied by randomized groups, and they also used that as their reasoning behind having randomized groups. The instructor stated: “But that way, they know it’s random, and if they hate their group, they know it’s not going to be a group forever.” The lack of connection to H. Group works well is even more interesting considering that literature suggests that randomized grouping’s main drawback is potential lack of group cohesion, along with the inability to equitably distribute underrepresented individuals [14,15].

Of the four instructors to discuss mixed-level groups and the two instructors to discuss same-level groups, only one instructor discussed both. The two instructors who discussed randomized groups discussed neither mixed-level nor same-level groups. This could imply that instructors have a preconceived notion about how to group their students. It could also explain instructors’ satisfaction in using same-level and randomized groups even though they are not supported by research as much as evenly mixed groups. The lack of grouping standards for the Next Gen PET curriculum allows for instructors to use those preconceived notions to make choices in their instructional practices. There are pros and cons to this, as instructors may be more comfortable with specific grouping methods in their classrooms, but that does not necessarily mean that those grouping methods are the most effective.

B. Shuffling groups

Shuffling Groups has one central node, AM. Shuffled groups; all but two nodes are direct causes or effects of instructors shuffling groups. The nodes with the next highest number of statements are AK. Concerns about student pacing and AJ. Good point in curriculum. Other than AJ. Good point in curriculum, all the causes of AM. Shuffled groups are related to students. Three are related to the instructors’ perception of the students, AK. Concerns about student pacing, AI. Difficult student, and AA. Students struggle. The last is AH. Student request, which is not related to the instructors’ perception of the students but could be considered as an instructor not perceiving an issue, causing the student to bring it to the instructors’ attention.

The effects of AM. Shuffled groups are also majority student oriented. Like the causes, we have nodes related to the instructors’ perception of the students, AC. Minimize student dissatisfaction, AF. Students build community and AD. Students on topic. Two of these, AC. Minimize student dissatisfaction and AF. Students build community, are related to students as well. The other, AD. Students on topic, is aimed at the instructors. Both AC. Minimize student dissatisfaction and AF. Students build community are strong central nodes with a large number of causes and effects, and their link to student satisfaction is important to instructors. The link to student dissatisfaction is an important counterpoint to the positive effects of shuffling groups.

FIG 2. Shuffling groups causal map. Nodes in this figure are lettered AA–AM., clockwise from the top, with AM, in the middle. The numbers in the parentheses denote (# of statements, # of faculty). Yellow square nodes are instructional practices, and indigo oval nodes are faculty reasoning. Color saturation of the nodes denotes its frequency in the dataset, with lighter colors having fewer statements.
Creating Groups Isolated RCMs

BA. Students are friends (1, 1)
BB. Students distracted (1, 1)
CA. Seating alphabetically (2, 1)
CB. Instructor learns about students (2, 1)
DA. Involve students in creating groups (2, 1)
DB. Student buy-in (2, 1)
EA. Small class size (2, 1)
EB. Students build community (2, 1)
FA. Physical space (2, 1)
FB. Instructor dissatisfaction (3, 2)
FC. Groups at different paces (1, 1)
FD. Few group combinations possible (2, 1)

FIG 3. Creating Groups isolated RCMs. Nodes in this figure are lettered top to bottom, left to right within each isolated RCM. A new first letter denotes a new map. The numbers in the parentheses denote (# of statements, # of faculty). Yellow square nodes are instructional practices, and indigo oval nodes are faculty reasoning. Color saturation of the nodes denotes its frequency in the dataset, with lighter colors having fewer statements.

dissatisfaction and AF. Students build community are related closely to the affect of the students, as with the causes AA. Students struggle and AI. Difficult student. This may suggest that instructors shuffle groups because they note a negative affect of their students, and they hope that by shuffling, they can see an uptick in positive affect of their students.

One isolated RCM was explicitly tied to the Shuffling Groups map. This RCM reads “If an instructor changes the curriculum, then they didn’t change groups, so the group works well.” Group works well and Changing the curriculum each had one statement from one instructor, while Didn’t change groups had two statements from two instructors.

Interestingly, the last of the isolated RCMs, one not explicitly tied to either map, reads “Instructors heard about a practice in their FOLC, so they tried that strategy.” Each node was found in two statements from two instructors. This causal statement is interesting because it shows that instructors are using the FOLC to field new instructional practices, suggesting that the FOLC was valuable as a learning community for these instructors.

IV. TAKEAWAYS

For instructors implementing the Next Gen PET curriculum, creating groups is a challenge. As can be seen from the RCMs, there are many variations in how faculty create groups, even though faculty have similar reasoning for using different grouping techniques. Ultimately, what faculty want are groups that work well where the students are not struggling.

While education research may have ideas about what grouping practices are the best, faculty may be staunch in their preconceived notions of grouping. This can make it difficult when curricula call for a particular grouping method, but faculty have their own modifications that they apply to the grouping methods.

A. Limitations and future work

This project has several limitations. The RCM method is frequently used on formal dialogue, such as company memos or political speeches [4]. Previous research in the physics education sphere has used RCM on instructor interviews [8]. Interview dialogue tends to be carefully thought out by the speaker, and interviewers are careful to not cut off or interject ideas into their interviewee’s speech, and can ask for the interviewee’s reasoning. The transcribed FOLC meetings used here are informal faculty discussions. This made identifying causal statements and their causes and effects difficult, particularly when the causal statements were spread across several turns of speech between different FOLC members.

Another limitation is the size of the dataset. This dataset encompasses a single cluster with thirteen FOLC members, and only eleven of those members are represented in the data presented in this paper. While this is a sufficient sized dataset for a case study on this particular FOLC, a larger dataset would be needed for any quantitative analysis or generalizations beyond this FOLC.

Future work on this project includes creating maps for the other facets of group work that were coded (group management, group assessment, and group activities,) and for whole class discussion. These maps will help Next Gen PET curriculum developers understand the various ways that faculty are implementing the curriculum. Additionally, RCM is intended to identify potential topics of future study. For example, future research could investigate why faculty seem to be staunch in their group formation methods or expand upon the lack of connection between Group works well and Randomized groups in the creating groups map, particularly considering the connection to student dissatisfaction. Does choosing randomized grouping indicate just a lack of time or more a lack of motivation to have cohesive groups? Future work could also include a comparison with other curricula that have had similar analysis done on faculty reasoning behind instructional practices.

ACKNOWLEDGMENTS

This work was funded by NSF DUE-1626496. Thanks to the Next Gen PET team for supporting this work.


Analyzing the dimensionality of the Energy and Momentum Conceptual Survey using Item Response Theory

Xian Wu and Matthew W. Guthrie
Department of Physics, University of Connecticut, Storrs, CT, 06269

Erin M. Scanlon
Department of Physics, University of Connecticut - Avery Point, Groton, CT, 06340

Yaoguang Li
AGQ Solutions, South Windsor, CT, 06074

There have been myriad conceptual inventories developed and used by the physics education research community (e.g., FCI, FMCE, BEMA, CSEM). However, continued examination of the psychometric properties of these inventories is important for improving inventory development and usage practices. In this study, we investigate the dimensionality of the Energy and Momentum Conceptual Survey (EMCS) test items and explore the structure of students’ understanding of these concepts (i.e., whether the student response data showed multiple distinct test traits). To investigate the psychometric properties of this survey, we surveyed a sample of 253 students participated in the pretest and 201 students for the posttest assessments. Statistical analyses, guided by Item Response Theory (IRT), were conducted using R software. The results of Bootstrap modified parallel analysis tests (BMPAT) revealed significant differences between unidimensional models and the actual data, indicating the presence of multidimensionality (i.e., dimensions correspond to the physics concepts/abilities being tested) in the EMCS test items. Exploratory analyses suggested that a 2 or 3-dimensional model best fits the data. However, categorizing items based on designated concepts did not improve the model fit. The findings imply that students may interpret the items differently than intended by the test designers.
I. INTRODUCTION AND BACKGROUND

Since its introduction nearly 40 years ago, the Force Concept Inventory (FCI) [1] has spurred numerous studies aimed at developing and analyzing research-based conceptual inventories which are generally used to measure student learning in physics courses [2-4]. In addition to many assessments focusing on topics other than forces, several alternative assessments to the FCI have been developed, including the Energy and Momentum Conceptual Survey (EMCS) [5]. The EMCS is a 25-item multiple-choice inventory specifically designed to evaluate students’ understanding of energy and momentum concepts in algebra-based and calculus-based introductory physics courses. This study examines data collected during the Fall 2022 and Spring 2023 semesters from a calculus-based Introductory Physics course designed for Introductory Physics for Life Science students (IPLS) and taught using the Studio physics approach. Our aim was to investigate students’ interpretations on the testing items and their understandings of energy and momentum concepts based on their responses. The analyses in this study were guided by Rasch Measurement Theory [6] for the alignment of the data (i.e., student responses) and the proposed statistical models (i.e., from one-parameter Item Response Theory). In the sections below, we will introduce the survey and the psychometric theories underpinning our analysis.

A. The Energy and Momentum Conceptual Survey

The 25 multiple-choice items in the EMCS were developed based on a study conducted at the University of Pittsburgh, which utilized free-response questions and think-aloud protocols [5]. The authors of the survey reported reasonable reliability coefficients [7] based on the results obtained from graduate and undergraduate students. Notably, numerous studies have acknowledged the soundness of the EMCS construction process [8-10] and the overall quality of the EMCS. However, to the authors’ knowledge, limited research has been conducted to investigate student performance on the EMCS.

B. Instructional Context

At the participating institution, the physics department implemented a Studio physics program for all calculus-based introductory physics courses on the main campus starting from the Fall 2019 semester. The Studio physics program primarily follows the SCALE-UP model [11], which promotes interactive and collaborative learning. For the course sequence offered to students in life science and pre-medical programs, the first author adopted an Introductory Physics for Life Sciences (IPLS) perspective, utilizing relevant resources available on the Living Physics Portal [12] for course design. The computer science program at the participating institution accepts a course sequence that is equivalent to the traditional engineering physics sequence. Consequently, the course sequence has generated consistent interest among computer science students.

II. METHODS

A. Item Response Theory Models

Item Response Theory (IRT) models a person’s response to a question as a result of both the person’s characteristics and the question’s properties [13]. In the unidimensional IRT framework, each student $\nu$ is associated with a proficiency parameter $\theta_\nu$, and each item $i$ is associated with a difficulty parameter $b_i$. The probability $P$ that student $\nu$ responds correctly to item $i$ (i.e., $x_{\nu i} = 1$) is expressed using a logistic function:

$$P(x_{\nu i} = 1) = \frac{\exp(\theta_\nu - b_i)}{1 + \exp(\theta_\nu - b_i)}. \quad (1)$$

Equation (1) represents the one-parameter logistic (1PL), also known as the Rasch model. In addition to the difficulty parameter $b_i$, the two-parameter (2PL) IRT model includes a discrimination parameter $a_i$:

$$P(x_{\nu i} = 1) = \frac{\exp(a_i(\theta_\nu - b_i))}{1 + \exp(a_i(\theta_\nu - b_i))}. \quad (2)$$

$a_i$ is a positive parameter that indicates an item’s ability to differentiate among students, although extremely high values are uncommon. The three-parameter (3PL) model further includes a pseudo-guessing parameter $c_{\nu i}$, which ranges between 0 and 0.5 and accounts for the chance of low-probability students guessing an item correctly:

$$P(x_{\nu i} = 1) = c_{\nu i} + (1 - c_{\nu i}) \frac{\exp(a_i(\theta_\nu - b_i))}{1 + \exp(a_i(\theta_\nu - b_i))}. \quad (3)$$

However, it is reasonable to acknowledge that a single physics testing item may encompass multiple concepts. Hence, student test results may not always conform to the assumptions of unidimensional IRT (i.e., a student can display multiple values for proficiency on one testing item). The compensatory multidimensional IRT model is commonly employed to address this concern by allowing for multidimensional proficiency and discrimination vectors for students $\theta_\nu$ and items $a_i$, while assuming a common difficulty parameter $b_i$ for all concepts. The updated equation is:

$$P(x_{\nu i} = 1) = c_{\nu i} + (1 - c_{\nu i}) \frac{\exp(a_i(\theta_\nu - b_i))}{1 + \exp(a_i(\theta_\nu - b_i))}. \quad (4)$$

In this study, instead of assuming that the data perfectly fits a specific IRT model, we examined the model to identify deviations from these assumptions [14]. One crucial aspect is assessing whether the data are unidimensional. We employed the Bootstrap modified parallel analysis test (BMPAT) to examine the unidimensionality of the data [15, 16]. BMPAT is a statistical technique that evaluates the underlying dimensions of multiple-choice questions by examining the relationships between the dichotomously scored items.

B. Model Fit Indices

Several model fit statistical procedures and criteria are available to compare competing models. In this study, we employed the Akaike information criterion (AIC), Bayesian information criterion (BIC), and the adjusted version of the BIC
indices mean better model fit. The AIC is calculated using estimates of the parameters. Generally, smaller values of these indices enable comparison of model-data fits our data. These indices enable comparison of model-data performance in studies involving a large number of parameters or small sample sizes [19, 20].

In addition to the likelihood function \( L \) and the number of estimated parameters \( p \), the BIC incorporates the sample size \( N \) in Eq. (6). The BIC is a more conservative estimate as it penalizes the selection of complex models [17]. The SABIC [18], on the other hand, uses \( N^* \) calculated by the equation:

\[
N^* = \frac{N + 2}{24}
\]

instead of \( N \). This adjustment has shown improved performance in studies involving a large number of parameters or small sample sizes [19, 20].

C. Data Collection
The data for this study were collected during the Fall 2022 and Spring 2023 semesters. The EMCS was administered along with a custom group work survey via Qualtrics to the IMLS course described in the background section. At the end of the survey, we added questions about the participants’ demographics including gender, race/ethnicity, and disability status. Additionally, an attention check question was included among the items of the EMCS to allow us to identify if a participant was not reading the questions. Students completed the survey twice during the semester: once as a pretest in the first week of the semester and once as a posttest in the last week of the semester. Participation in the survey was anonymous, and at the end of the survey, students were requested to generate an identification code. The identification code contained five letters and four numbers from four pieces of information about the student and their major family events. The survey was announced as an extra credit assignment, and all students in the class were eligible to receive 0.5% of their course grade as extra credit if the response rate reached 80% or higher. A total of 454 students participated in the survey and passed the attention check, with 253 students completing the pretest and 201 students completing the posttest. 2 students failed the attention check and their data were removed from the data set being used in the present study.

D. Analysis
The statistical analysis in this study was conducted using R statistical software [21]. We utilized two R packages: ltm [22] for the unidimensional checks, and mirt [23] for the analyses related to multidimensional IRT.

The research question guiding this study was: What is the optimal statistical model for analyzing the structure of the test responses obtained from the EMCS? Considering the anticipated inclusion of more data in the coming academic years, the results and conclusions presented in the following sections are exploratory in nature.

III. RESULTS
A. Unidimensional Fit Analyses
To explore the dimensionality of the data, Bootstrap modified parallel analysis test (BMPAT) was conducted on the combined pretest and posttest results, as well as separately on the pretest results and posttest results. The unidimensional 1PL, 2PL, and 3PL models were tested, and the results of all nine tests are presented in Fig. 1. The vertical axes all show eigenvalue while the horizontal axes all show eigenvalue number. The black lines with circles represent eigenvalues calculated from the real data and the red lines with triangles represent eigenvalues calculated from the average simulated data. The average simulated data was generated by repeating the Monte Carlo simulations 100 times while assuming unidimensionality. The \( p \)-value was calculated for comparing the second eigenvalues of the real data and simulated data. The equation is:

\[
p = \frac{n + 1}{B + 1}.
\]

where \( n \) is the number of times when the simulated second eigenvalue is larger than the one from the real data and \( B \) is the number of the Monte Carlo simulations being repeated [22]. All tests indicated a statistically significant difference between the simulated and the real data with \( p = 0.0099 \).

Because \( p = 0.0099 \) in every test, we can reliably conclude that none of the simulated second eigenvalues are larger than the one from the real data. Therefore, the null hypothesis, that the data follow a unidimensional IRT model, can be rejected. This suggests that none of the tested models (combined, pretest, or posttest) adequately fit the data, indicating the presence of multidimensionality in the EMCS test items. The number of eigenvalues shown in Fig. 1 corresponds to the number of dimensions. The second eigenvalue values in the observed data were consistently higher than the simulated values across all model comparisons.

B. Multidimensionality fit Analysis
These findings prompted exploratory analyses to determine the optimal number of dimensions required for the model. A series of 3PL models ranging from 1 to 5 dimensions were examined, and the model fit statistics are presented in Table I. The AIC value decreased until the 4-dimensional model, while the BIC value decreased until the 2-dimensional model. The SABIC value, on the other hand, decreased until the 2-dimensional model.
FIG. 1. Parallel analyses of the 1PL, 2PL, and 3PL models with stacking pre- and posttest data together and pre- or posttest data only. The vertical axes all show eigenvalue while the horizontal axes are all show eigenvalue number. The second eigenvalues in the observed data and simulations are listed, along with the corresponding $p$ value.

Since SABIC and BIC tests are more conservative than AIC test, the optimal model may contain 2 or 3 dimensions. Considering the design of the EMCS, which assigned items 1, 2, 4, 6, 8, 9, 12, 13, 15, 17, 20, 22, 24, and 25 to energy and items 3, 5, 7, 10, 12, 14, 18, 19, 21, and 23 to momentum concepts with item 16 for both energy and momentum concepts, we explored three candidate models: two 2-dimensional 3PL models, where item 16 was assigned to either the energy or momentum group, and one 3-dimensional 3PL model, where item 16 formed a standalone third group. Since we would like to test whether the models with extra dimensionalities can align with the data better than the unidimensional model, another round of exploratory analyses was conducted to compare these candidate models against the 1-dimensional 3PL model as the baseline. The model fit statistics can be found in Table II. Since the 1-dimensional 3PL model has the smallest AIC, SABIC, and BIC values among all models listed in Table II, none of the candidate models demonstrated an advantage over the 1-dimensional 3PL model. The statistical evidence suggests that categorizing the EMCS items based on the designated testing concepts does not yield a better model compared to grouping all items together. Given the data of test responses reveals how students interpret the items on the test, our results showed students did not recognize the underlying concepts of items the same way as how the items were designed.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>AIC</th>
<th>SABIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13522.50</td>
<td>13593.34</td>
<td>13831.36</td>
</tr>
<tr>
<td>2</td>
<td>13401.31</td>
<td>13494.81</td>
<td>13809.00</td>
</tr>
<tr>
<td>3</td>
<td>13390.17</td>
<td>13505.39</td>
<td>13892.58</td>
</tr>
<tr>
<td>4</td>
<td>13352.11</td>
<td>13488.11</td>
<td>13945.12</td>
</tr>
<tr>
<td>5</td>
<td>13382.53</td>
<td>13538.36</td>
<td>14062.01</td>
</tr>
</tbody>
</table>
TABLE II. MIRT model fit statistics with different ways of grouping. “1D” is the 1-dimensional 3PL model. “2D16M” is the 2-dimensional 3PL model with assigning item 16 to the momentum group. “2D16E” is the 2-dimensional 3PL model with assigning item 16 to the energy group. “3D16ME” is the 3-dimensional 3PL model with item 16 being in a third group alone.

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>SABIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D</td>
<td>13522.50</td>
<td>13593.34</td>
<td>13831.36</td>
</tr>
<tr>
<td>2D16M</td>
<td>13727.73</td>
<td>13798.56</td>
<td>14036.58</td>
</tr>
<tr>
<td>2D16E</td>
<td>13723.83</td>
<td>13794.67</td>
<td>14032.69</td>
</tr>
<tr>
<td>3D16ME</td>
<td>13749.74</td>
<td>13820.57</td>
<td>14058.59</td>
</tr>
</tbody>
</table>

IV. DISCUSSION

The dimensionality of conceptual inventory testing data is influenced by the interplay between how each item is designed and how students interpret and respond to those items. It is important to note that different student populations and institutions may yield varying results in this discussion. For instance, a study conducted by Wang and Bao [24] analyzed data from approximately two thousand students taking the FCI at the Ohio State University over a span of five years. In their study, both the pretest and posttest data aligned well with a unidimensional 3PL model. On the other hand, a study by Stewart et al., [25] involving over four thousand students from a different institution, suggested an optimal model with nine dimensions. Furthermore, other studies have suggested 5 or 6-dimensional models [26, 27].

In the present study, the analyses demonstrated the need for a multidimensional model to effectively explain the EMCS testing results. However, categorizing items based on the intentional testing concepts did not result in a better model compared to the unidimensional ones. This suggests that students may not interpret the items in the same way as intended by the test designers, and highlights the complexity of student understanding, as well as the need to further investigate the factors influencing item interpretation and response patterns.

When considering the use of multidimensional IRT, sample size becomes a crucial concern. Many of the previously cited works have employed samples with thousands of participants. However, according to the guidelines provided by Linacre, a sample size of around 150 participants can still achieve a high level of confidence, typically around 99% in most situations [28]. Given the exploratory nature of this study, the current sample size is deemed sufficient.

V. IMPLICATIONS

Based on the study and its conclusion, several implications can be suggested for researchers and educators. Firstly, it is important for researchers to carefully examine the structure of conceptual inventories and consider multidimensionality when analyzing and interpreting the results. This can provide a more accurate understanding of students’ conceptual understanding and guide the development of more effective assessment tools.

Secondly, the findings emphasize the need for educators to be aware that students may interpret conceptual inventory items differently than intended. It is therefore critical to highlight the importance of aligning instruction and assessment to bridge any gaps in conceptual understanding and promote more accurate interpretations of the concepts being assessed.

Overall, this study underscores the ongoing need for rigorous examination of conceptual inventories and a deeper understanding of how students engage with conceptual inventory items. By incorporating these insights into research and practice, educators can better support students’ conceptual learning and improve the effectiveness of conceptual assessments.

VI. FUTURE WORK

The present study opens up two potential directions for future research. Firstly, as we continue to accumulate data from more participants over multiple semesters, the MIRT analyses will gain increased statistical power, enabling more definitive conclusions regarding the structure of the testing items and how students comprehend the relevant concepts. With larger sample sizes, we will be better equipped to explore the multidimensionality of the EMCS test items and gain a deeper understanding of students’ conceptual understanding.

Secondly, we plan to investigate how demographic factors influence student conceptual learning outcomes. Starting from the Fall 2022 semester, we have been collecting demographic information, including gender, ethnicity, and disability, potentially enabling us to examine the multifaceted fairness of the EMCS items and our IPLS course. By analyzing the data in conjunction with demographic factors, we can assess potential disparities in student performance. This research endeavor will provide insights into how demographic characteristics may interact with student understanding of energy and momentum concepts.

ACKNOWLEDGMENTS

This work is partially supported by the internal grant from University of Connecticut’s College of Liberal Arts and Sciences.
[17] S. I. Vrieze, Model selection and psychological theory: A discussion of the differences between the akaike information criterion (AIC) and the bayesian information criterion (BIC), Psychological Methods 17, 228 (2012).
Students’ use of symmetry as a tool for sensemaking

Tamara G. Young
Department of Physics and Astronomy, University of Utah, 115 S. 1400 E. 201 JFB, SLC, UT 84112

Lauren Barth-Cohen
Department of Educational Psychology, University of Utah,
1721 Campus Center Dr. SAEC 3220, SLC, UT 84112
Department of Physics and Astronomy, University of Utah, 115 S. 1400 E. 201 JFB, SLC, UT 84112

Sarah K. Braden
School of Teacher Education and Leadership, Utah State University, 2805 Old Main Hill, Logan, UT 84322

Symmetry is a particularly important sensemaking tool in physics; physicists use symmetry for everything from classifying crystalline solids to deriving the foundational laws of physics. In addition, symmetry is embedded in the Next Generation Science Standards (NGSS) crosscutting concepts (CCCs). In this research, we explore how students use symmetry in ways that are consistent with how physicists use symmetry. We present 5 cases of students engaged in an NGSS-aligned lesson on creating models of magnetism. In each case, we highlight one of the ways that we have found students using symmetry as a sensemaking tool: to describe, classify, predict, explain, and solve problems. These results contribute to conversations on how students learn to “think like a physicist” and how students engage in the NGSS crosscutting concepts and science and engineering practices (SEPs).
I. INTRODUCTION AND BACKGROUND

Exploring how students learn to think like physicists is important in Physics Education Research (PER). One tool that physicists use to make sense of the universe is symmetry. Symmetry has been described as the foundational issue in deriving the laws of nature [1, 2]. Symmetry played an important role in such laws as energy conservation, Maxwell’s equations, and General Relativity [2, 3]. Symmetry is used as a method of classifying phenomena and in predicting the properties of new phenomena [3], for example, in crystallography [4], solid state and molecular physics [5], and in high energy and particle physics [6]. Physicists use symmetry to explain phenomena. For example, Wigner used the rotational symmetry of allowed electron orbitals to explain atomic spectra [1]. Physicists use symmetry in simplifying and solving problems; many integrals in undergraduate electrodynamics textbooks are solved “by symmetry” [e.g., 7]. The use of symmetry in physics is profoundly important.

Given the importance of symmetry for physicists along with the PER goal of exploring how students learn to think like a physicist, it becomes relevant to consider how students reason with symmetry. Furthermore, symmetry is a topic that is understudied in PER. Notably, there is some work on undergraduate students in electromagnetism courses that either focus on their use of symmetry [e.g., 8–10] or include students’ uses of symmetry in the analysis [e.g., 11, 12]. However, this work primarily focused on students’ difficulty with using symmetry in solving problems in electromagnetism. Demonstrations of students’ productive uses of symmetry are limited in this work.

Examining students’ reasoning with symmetry in PER is also important because the notion of symmetry is embedded in the Next Generation Science Standards (NGSS), a nationwide set of standards that many states in the U.S. have adopted) crosscutting concepts (CCCs) of patterns; and scale, proportion, and quantity [13]. The Framework for K-12 Science Education describes the CCC of patterns in this way.

Patterns. Observed patterns of forms and events guide organization and classification, and they prompt questions about relationships and the factors that influence them([14], p. 84).

In the NGSS, the discussion of patterns mentions symmetry as one type of pattern that can “guide organization and classification” and “prompt questions” about phenomena. It also points teachers to resources on how symmetry is used in physics and other sciences [15]. In scaling symmetry, the phenomena are the same at different scales. This type of symmetry is embedded in the CCC of scale, proportion, and quantity [13]. Within this CCC, there is recognition that phenomena vary across scales, which can be interpreted as a version of scaling symmetry. Additionally, many ways that physicists use symmetry align with the NGSS science and engineering practices (SEPs) [13]. For example, in constructing explanations and designing solutions or analyzing and interpreting data. Given symmetry’s role in NGSS, and the growing recognition of the importance of CCCs in science learning and teaching [16, 17], we suggest there is a need for greater emphasis on symmetry moving forward.

This leads to our research question: how do secondary students use symmetry in ways that are similar to how physicists use symmetry when engaging in NGSS-aligned lessons?

II. THEORETICAL FRAMEWORK

This research is situated in resources theory, particularly Knowledge in Pieces (KiP). In this framework, students construct knowledge using intuitive knowledge resources that they gain from interacting with their environment [18, 19]. Specifically relevant to this study, symmetry resources may be one type of intuitive resource [18], but little work has been done to determine the nature of these resources. While backgrounded in this study, the KiP framework guided the design of the interview and instruction, which aimed to pull out student resources regarding symmetry.

Additionally, we are interested in the ways that students use symmetry in sensemaking. Sensemaking is a process in which students will create or revise explanations in order to “figure something out” [20]. Here we are interested in the ways students use symmetry in figuring things out. For example, in describing things, classifying things, explaining things, predicting things, or solving problems.

III. METHODS

A. Data acquisition

To explore how secondary students use symmetry in ways that are similar to physicists, we used a lesson on the NGSS disciplinary core idea of forces and interactions [13], in particular, the forces and interactions of magnetism. We selected magnetism to explore student use of symmetry due to the types of symmetry present in a standard dipole magnet. A dipole magnet has reflective and rotational symmetry in the magnetic field and the orientation of the poles. In addition, a permanent magnet has scaling symmetry in the magnetic domains and translational symmetry in the arrangement of these domains. Students participating in this research engaged in an NGSS-aligned lesson in which they constructed models of magnetism by interacting with magnetic phenomena and drawing models of what was happening inside and around the magnets [21].

This research was conducted in a large metropolitan area in the Intermountain West. Participants were recruited in two ways. The participants for Group A were involved in an afterschool program for refugee students that focused on Science, Technology, Engineering, Art, and Math, (STEAM). These students engaged in a lesson on magnetism as part of their afterschool program; later, they participated in small-group
interviews with the researchers and colleagues. The data was taken from the interviews. In total, 13 students in grades 7-12 were interviewed in 5 groups. Each interview lasted between 1-1.5 hours. The participants for Group B were recruited through social media and snowballing techniques, in which volunteers were asked to identify friends who might also be interested. These participants were interviewed by the lead author in small-group interviews in their self-identified small friend groups. For Group B, 11 students in grades 7-9 were interviewed in 5 groups. There are differences in the experiences the students had with formal magnetism instruction both between and within the groups.

In all the interviews, the students explored magnetic phenomena and then constructed models of magnetism based on their observations. There were four parts to the interviews. First, the students were introduced to the motivating phenomenon, a floating paperclip. Next, they interacted with magnet stations that we developed to allow the students to make important observations about magnets (see Table I). Next, the students were asked to draw what was happening inside and around the magnets. During this step, the interviewer asked the students questions intended to push them to construct a model that explained the observed phenomena. Finally, the interviewers asked students questions about the model and about running the model. The interview protocol included questions designed to elicit the use of symmetry in constructing and describing the models. For example, to elicit the use of rotational symmetry, the interviewers asked the students about how the shape of the magnetic field would change if they rotated the magnet in different ways. Cameras were set up to record the heads and torsos of the students as well as capture the written artifacts as the students worked on their models. After the interviews, the models were collected, and the interviewers provided written or verbal field notes. We also collected backup audio during the interviews.

<table>
<thead>
<tr>
<th>Magnet Station</th>
<th>Description/Image</th>
<th>Intended Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Floating paperclip</td>
<td><img src="https://example.com/clip.png" alt="Image" /></td>
<td>This is the anchoring phenomenon. The students are asked what is happening in and around the magnet and paperclip to cause it to float.</td>
</tr>
<tr>
<td>2. Bouncing magnets</td>
<td><img src="https://example.com/bounce.png" alt="Image" /></td>
<td>Magnets have different poles.</td>
</tr>
<tr>
<td>3. Broken magnets</td>
<td><img src="https://example.com/broken.png" alt="Image" /></td>
<td>When a magnet is broken, it forms two poles. The way the magnet is broken affects the behavior of the broken magnet.</td>
</tr>
<tr>
<td>4. Magnetic viewer</td>
<td><img src="https://example.com/viewer.png" alt="Image" /></td>
<td>The shape of the magnetic field. The magnetic field looks different for a disk magnet than for a bar magnet.</td>
</tr>
</tbody>
</table>

We then coded these segments based on the types of symmetry the students appeared to be using: rotational symmetry, reflective symmetry, scaling symmetry, translational symmetry, and pseudovector symmetry. We also noted the context in which the students were using symmetry. For example, were the students discussing a single magnet or several magnets? Were the students discussing the particles that make up the magnet? etc. After this initial round of coding, we selected exemplar cases and added non-verbal elements (e.g., gestures, positions) to the transcripts to aid interpretation.

We then completed a second round of coding informed by our theoretical framework of sensemaking, which was based on how students used symmetry. The codes for symmetry that emerged from the data include: describing phenomena, classifying phenomena, explaining phenomena, making predictions, and solving problems. From this round of coding, we selected cases (from the original exemplars) that highlight how students used symmetry similarly to physicists [25].

IV. ANALYSIS

Here we present cases of the five ways we observed students using symmetry in our interviews (see Table II); we provide more in-depth descriptions for three cases that align strongly with the NGSS SEPs. In these cases, we find that students use symmetry as a sensemaking tool. They use different kinds of symmetry, and they use symmetry for different purposes to meet specific goals. All of the symmetry types discussed in the following cases are embedded in the NGSS CCC of patterns.

A. Case 3: Using Scaling Symmetry to Generate an Explanation

In addition to the NGSS CCC of patterns, scaling symmetry is also embedded in the NGSS CCC of scale, proportion, and quantity. An exemplar use of scaling symmetry to explain a phenomenon occurred in an interview with two stu-
TABLE II. Students’ use of Symmetry.

<table>
<thead>
<tr>
<th>Use of Symmetry</th>
<th>Context</th>
<th>Transcript</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Using Bilateral Symmetry to Describe</td>
<td>Raj uses gestures to indicate the symmetry in the magnetic field to describe its shape (magnetic station #1).</td>
<td>“So, it goes on the opposite side of the, kind of like this. So this is on top and bottom &lt;drawing a line back and forth with his hands&gt;, I couldn’t really show because it’s a 2-D graph, but like it’d be like, like top to bottom &lt;drawing a bilateral magnetic field in the air&gt;.”</td>
</tr>
<tr>
<td>2. Using Rotational Symmetry to Classify</td>
<td>Wesley uses the symmetry of the broken magnets (magnetic station #3) to classify a broken donut magnet (magnetic station #2) as a disk magnet, not a bar magnet.</td>
<td>“So you can see they stick just regularly &lt;sticks pieces of the magnet together&gt; and if we rotate it, &lt;turns the magnet over&gt; it does not care &lt;sticks pieces of the magnet together again&gt;.”</td>
</tr>
<tr>
<td>3. Using Scaling Symmetry to Generate an Explanation</td>
<td>Raj and Habte use the symmetry between the parts and the whole to explain why the paperclip sticks to the magnet but the pencil does not (magnetic station #1).</td>
<td>How to represent the particles of the pencil. Habte: “Just little dots like that &lt;drawing little dots on the model&gt;.” Comparing this to the particles of the paperclip. Raj: “Maybe both &lt;draws pluses and minuses on the model&gt;.”</td>
</tr>
<tr>
<td>4. Using Rotational Symmetry to Make Predictions</td>
<td>Ali uses the rotational symmetry of a dipole magnet to predict that the magnetic field of a disk magnet will look similar to a bar magnet if rotated 90 degrees (magnetic station #4).</td>
<td>Observing a disk magnet. “It’s [the iron filings] all getting attracted to that side because... I bet if you turned it over this way &lt;gesturing to rotate the magnet 90 degrees&gt;, it would look the same as that &lt;indicating the bar magnet&gt;.”</td>
</tr>
<tr>
<td>5. Using Bilateral Symmetry to Solve a Problem</td>
<td>Andrew uses the magnet’s bilateral symmetry to figure out why the broken disk magnet will not go back together (magnetic station #3).</td>
<td>“So if this is north, and the bottom’s south, so then, I break it apart, now these two are north, so they don’t like each other and these two are, and the bottoms are south, so they don’t like each other. So it pushes away. But if you flip it around like before. And then they’ll stick. It makes sense now. &lt;throws arms wide and pushes chair away from table&gt;.”</td>
</tr>
</tbody>
</table>

dents in Group A who were trying to figure out why the paperclip will stick to the magnet but is not magnetic itself (magnetic station #1). In this case, Raj and Habte, both 11th-grade students, had constructed models in which the magnet was a dipole composed of positive and negative particles, with more negative particles on one end and more positive particles on the other end. This model represents an instance of imperfect scaling symmetry in that the pieces that make up the magnets are approximately smaller versions of the magnet.

Raj wants to discuss the paperclip because he is unsure how to draw it since it “attracts to both sides” of the magnet. Tamara asks Raj and Habte to think about the difference between the particles that make up the paperclip and the particles that make up something non-magnetic like the wooden part of a pencil. The students decide that while both the paperclip and the wooden pencil are magnetically neutral, Habte states that the particles of wood should be represented by “just little dots like that <drawing the little dots on the model>,” and Raj agrees by saying “no charges.” This model uses scaling symmetry in that both the pencil and the particles that make up the pencil are non-magnetic. Tamara then asks the students what the particles of the paperclip would look like, and Raj responds with “maybe both” charges and then draws the particles with a mix of positive and negative particles. When they subsequently ran the model, the students decide that the charges of the paperclip will be “more clustered” when in the magnetic field, allowing them to attract to either side of the magnet, but not be a magnet themselves. Thus, their models explain the difference between magnets, magnetic materials, and non-magnetic materials. This is important in that explaining phenomena is both in the NGSS standards and one of the ways that physicists use symmetry.

B. Case 4: Using Rotational Symmetry to Make Predictions

In our data, we find that the students are able to use their models to make predictions, which is one of the goals of the NGSS SEP developing and using models. Additionally, even before they develop their models, they are able to use the symmetry of the system to make predictions. Here we see students using the rotational symmetry of a dipole magnet to predict that the magnetic field of a disk magnet will look similar to the magnetic field of a bar magnet if rotated 90 degrees.

In this case, during an interview with two students in Group A, Tamara is showing Ali, an 11-grader, and Raphael, a 10th grader, the magnetic field through the magnetic field viewer (magnetic station #4). Tamara first shows the students the bar magnet, and the students note how the shape sort of balloons out. Tamara asks them to predict the shape of the disk magnet. Raphael says he thinks the shape would look “the same, just stronger, probably.” However, when they try it, Ali
notes that “it’s all getting attracted to that side because... I bet if you turned it over this way, it would look the same as that.” Here Ali is using the rotational symmetry of the magnetic field to correctly predict that the fields of the bar and disk magnets are the same if rotated 90 degrees. After Ali makes this prediction, Tamara turns the disk magnet on its side, and they observe that while it doesn’t look identical, it has the same type of ballooning shape that the bar magnet did. This is noteworthy because making predictions is embedded in the NGSS standards and is an important way physicists use symmetry.

C. Case 5: Using Bilateral Symmetry to Solve a Problem

One problem of particular interest in this study was the problem of why the broken magnet will not fit back together. While not every group of students was able to solve this problem, many were, and they all used the symmetry of the magnet to figure this out. Here we present how Andrew, a 7th-grade student from Group B, used the magnet’s rotational symmetry to figure out why the broken disk magnet will not go back together (magnetic station #3).

In this case, Tamara is asking the students to run their models. Tamara asks Andrew what would happen with his model if they broke it across the poles. Andrew recognizes from the symmetry of the model, that “the tops, I think, would stay north and the bottoms would stay south.” Tamara agrees and asks what would happen to the pieces. Andrew reasons that the two north sides would repel and the two south sides would repel. But if you turn one of the sides over, they will stick. At this point, it is clear from Andrew’s enthusiastic “it makes sense now! <throws arms wide and pushes chair away from table>” Andrew is pleased that he has solved the problem of the broken disk magnet. This is meaningful because solving problems “by symmetry” is an experience that every physics student is familiar with; physicists use symmetry as a problem-solving tool in both teaching and research.

V. DISCUSSION

In this paper, we observed students using symmetry as a sensemaking tool in describing phenomena, classifying phenomena, generating explanations, making predictions, and solving problems. Here the students are in the process of learning to think like physicists. While this analysis may not have captured this learning, it provides insight into the process, and future work may be able to capture it.

The students’ conclusions about symmetry are not what an expert would say; they do not use the same language as a physicist. Rather, the students use language that is familiar and comfortable to them to describe the magnetic phenomena, which is frequently not the vocabulary that a scientist would use. Additionally, they do not discuss the concept of symmetry; rather, they express symmetry-related concepts through their embodied actions and tone of voice that indicates a sense of obviousness to symmetry-related questions. Nevertheless, the students have symmetry resources that are productive and can be leveraged for future learning.

Our research primarily focused on geometric symmetry, given that this type of symmetry is highlighted in the physical aspects of permanent magnets. The cases presented here are all instances of the geometric symmetry of the magnet and other items. Case 2 and Case 4 involve executable types of symmetry (i.e. rotations and translations). Case 3 uses non-executable types of symmetry (i.e. reflections and scaling). There is some ambiguity in Case 1 and Case 5, in that bilateral symmetry can occur as a result of both rotation and reflection. While students most often used symmetry when asked triggering questions, occasionally, the students would use symmetry spontaneously. These spontaneous uses of symmetry occurred most often when the students were engaged in productive struggle during their sensemaking process. That there are other types of symmetry, and other uses of symmetry, than the ones we focused on in our research, leads us to ask if students also use other types of symmetry in sensemaking, such as the translational and rotational symmetry of spacetime. Or how might we trigger the use of symmetry related to formulating mathematical laws of physics?

An additional contribution of this study is in instructional design, curriculum, and teaching. The instruction was built to forefront the NGSS science and engineering practice of developing and using models. However, when looking at the data, the students are using other SEPs, such as constructing explanations and analyzing and interpreting data. Additionally, in constructing their models, the students are using the CCC of cause and effect in addition to the CCCs of patterns; and scale, proportion, and quantity. In as much as making the SEPs and CCCs explicit in the lesson is a best practice when teaching NGSS-aligned lessons [16, 17], we suggest teachers and instructional designers working with NGSS-aligned lessons revisit the lessons to look for additional SEPs and CCCs that may be hidden in the lesson and bring these concepts to the forefront. In particular, we encourage teachers and instructional designers to be more explicit about symmetry in their NGSS curriculum and teaching.

ACKNOWLEDGMENTS

This research is sponsored in part by NSF Award 2005973. Thanks to Jason May, Sara Gailey, Adrian Adams, Laurel Dias, and Rachel Francom for help with interviewing. Thanks to our collaborators for help with data collection: Tino Nyawelo, Jordan Gerton, Melanie Valera Garcia, Raquel Goldrup, Bolaji Bamidele, Sterling Johnson, Joseph Kifflon, and all those who make the afterschool program possible. Thanks to all the students, their parents, and teachers who supported this research. Special thanks to Chris Haas for their critical lens.
Analyzing Physics Majors’ Specialization Low Interest Using Social Cognitive Career Theory

Dina Zohrabi Alaee (she/her), Keegan Shea Tonry (they/them), and Benjamin M. Zwickl (he/him)

Physics Department, Rochester Institute of Technology, 84 Lomb Memorial Dr, Rochester, NY, 14623-5604

As students pursue a bachelor’s degree in physics, they may ponder over which area to specialize in, such as theory, computation, or experiment. Often students develop preferences and dislikes, but it’s unclear when this preference solidifies during their undergraduate experiences. To better understand, we interviewed eighteen physics majors at different stages of their degrees regarding their interest in theory, computation, and experimental methods. Out of the eighteen students, we analyzed only nine students who rated computation and theory the lowest. Our analysis did not include interest in the experiment because the ratings were less negative. We used Social Cognitive Career Theory (SCCT) and Lucidchart to analyze students’ responses and create individual graphical representations of the influences for each student. Through this, we uncovered how various factors such as learning experiences, self-efficacy, and outcome expectations influenced their low interest in a particular method. We found that lack of knowledge and experience is often the main reason why self-efficacy was lower. Students’ lack of interest is also influenced by negative outcome expectations (e.g., math-intensive and a bad work-life balance) more than other SCCT factors. Our findings could help physics departments and educators identify positive and negative factors that could lead to a more motivating and inclusive physics curriculum.
I. INTRODUCTION

In order to provide students with the best possible education, physics departments should prioritize creating a learning environment that is engaging and challenging and also emphasizes the importance of different method specializations such as theory, computation, and experimental skills. By doing so, students can develop a deeper interest in these areas and gain a better understanding of their future goals and their career options. Students’ actual experiences as physics majors may fall short of this goal. For example, some physics majors may not have access to much computational coursework [1], and most will get far more practice with theoretical problem-solving than with hands-on experimental work.

According to statistics from the American Institute of Physics [2], 50% of physics bachelor’s degree recipients who seek jobs in the private sector are working in computing and engineering-related jobs. For those who pursued graduate school, about 50% used experimental methods for their research, 37% did theoretical work, and 10% used observational and other methods [3]. However, there isn’t much research on what factors influence these career choices. Additionally, substantial prior work has focused on how students’ physics identity [4] and out-of-class learning experiences in high school affect their decision to study physics in college [5, 6]. Other work has studied interest in physics and long-term retention [7].

The focus of this study is to gain insight into the factors that affect the career transition of physics graduates. Due to a need for brevity, we focus on only negative interest formation. Understanding students’ low interest is important so we can address the underlying causes and create more inclusive environments that encourage diverse career exploration. The main research question addressed in this study is: What factors and experiences contribute to physics majors developing a low interest in two particular method specializations (theory and computation) in physics? By exploring this topic, we hope to better understand the decision-making processes of physics students and identify strategies that support rather than hinder their future career paths.

II. BACKGROUND

The theoretical framework for this study is based on Social Cognitive Career Theory (SCCT), which reveals different aspects of career and interest development [8, 9]. Figure 1 shows the SCCT map, which consists of self-efficacy, outcome expectations, proximal environment influences, interest, and learning experiences. In SCCT, learning experiences are any learning opportunity that may have had an influence, self-efficacy refers to belief in one’s abilities, outcome expectations are the perceived consequences of career choices, proximal environmental influences are external factors beyond the learning experiences, and interest reflects the degree of positive or negative sentiment and an intent to pursue in the future. These are the important blocks of the SCCT map that differ the most between students. We have also included the influential factor of a sense of belonging in our SCCT map [10–12]. Sense of belonging refers to individuals feeling a sense of fitting in and connection within a specific community. This feeling of belonging can drive motivation, influence the formation of interests, and ultimately impact their chosen career path.

When considering the development of interest in certain method specializations in physics, it is important to explore what factors and experiences may contribute to this process. SCCT posits that interests are shaped by self-efficacy and outcome expectations that are influenced by learning experiences. It is assumed that students gravitate towards fields and topics for which they have higher self-efficacy. Having poor performance in previous courses could lead to lower self-efficacy. On the other hand, outcome expectations are what a student expects work or life will be like as a consequence of working in a particular field or with a specific method specialization. SCCT assumes that people will engage in activities that they expect to be rewarding. One example of a negative outcome expectation is that a particular career has lots of pressure and a poor balance between work and life. Proximal environmental influences refer to the contextual variables that influence students’ interests, such as family or peer support that can impact a student’s interest.

Finally, following along the map, choice goals and choice actions refer to an intent or action to take part in activities related to the method of interest and achieve a certain level of success. In our study about low interests, a choice goal or choice action could be avoiding a particular method. Through the exploration of this topic, the study hopes to gain insight into the low-interest formation experienced by physics undergraduate students.

III. METHODOLOGY

Our study involved semi-structured interviews with undergraduate physics students in the Summer 2022 and Fall 2022. We recruited participants for our study through various channels, including a student-organized Discord server for physics majors, an email through the physics department, and visit-
ing a course for first-year physics majors. All participants were from the same large private research university in the United States, which has about 200 undergraduate physics majors. In total, we interviewed eighteen physics students who were at different stages of their studies. Out of the 18 physics students, we chose to analyze the five lowest ratings for computation and the five lowest ratings for theory, which included nine distinct students. Since the overall rating of the experimental method was high, we did not include it in our analysis. Of the 9 participants, 4 described their gender as “man,” 4 as “non-binary,” and 1 as “woman”. Regarding race and/or ethnicity, 7 described themselves as “White”, 1 as “Latino”, and 1 as “Asian”. Finally, 4 were in their fourth year, 3 in their third year, and 2 in their first year. Participation incentives were offered in the form of a $15 gift card. During these interviews, the interviewer started by asking the students to explain each method to ensure that they fully grasped the concepts and provided clarifications as necessary. Then, the interviewer directly asked students to rate their level of interest in each method on a scale of 0 to 10, with 0 indicating no interest at all and 10 indicating very high interest. Throughout the paper, we use a shorthand notation to indicate interest levels, such as (T-3) to mean a theory interest rating of 3 or (C-5) to mean a computation interest rating of 5. We delved deeper by asking follow-up questions tied to SCCT factors to uncover the various influences on their level of interest. Each interview was recorded, transcribed using Otter.ai, and transcription errors were corrected prior to analysis. To investigate the formation of student’s interests, we were inspired by the phenomenography method [13, 14] to understand the variation in experiences that led to a low interest in theory or computation. The limited sample size of our study limits our ability to reach saturation in data collection and data analysis. First, for each student, we created an individual diagram of each student’s influences using Lucidchart [15] (see Fig. 2-a). Each node in the diagram was a short quotation or summary about factors that influenced their interest, and links were based on causal connections in the transcript. After creating 10 separate maps, we coded each student’s statements into SCCT categories such as learning experiences, self-efficacy, outcome expectations, as well as a sense of belonging (priori coding). Then, we merged individual maps by combining similar quotes and created one comprehensive map for theoretical interest and one for computational interest (see Fig. 2-b). This allowed us to see the variation across students’ experiences and highlighted similar and contrasting factors that impacted their interest formation. Finally, for each method, emergent subcategories (inductive coding) were identified within the SCCT and sense of belonging constructs (see Fig. 2-c).

IV. RESULTS

A. Interest Levels

We characterized interest based on students’ numerical ratings and their qualitative comments. Based on numerical ratings, students in our study had a more negative perception of theory (T) (ratings: 1,2,2,2,4; Av. = 2.2) compared to the computation (C) (ratings: 3,5,5,5,5; Av. = 4.2). For the other participants who were not analyzed here, the average rating was 7.4 for theory (ranging from 6-10) and 7.0 for computation (ranging from 5-10). The average for experimental work among 18 students was 8 (ranging from 5-10).

Students’ interest was also indicated by their language when describing a particular method. Some students expressed strong negative emotions and language when discussing their lack of interest in theory. For example, Skyler, who rated theory a 2 (T-2), stated, “I hated the materials. It wasn’t necessarily a negative class experience. I just did not enjoy the material in the slightest... Just cringe. The professors are all crazy, and I’m blanking on every class I’ve ever taken. Modern was garbage.” Skyler continued that they had a supportive professor that tried to help them, but they were not paying attention. They said, “I voluntarily excluded myself from theoretical areas and feel dread to get involved in this type of work.” Armond (T-1) also used strongly negative language, describing theory as “miserable,” “class was really just gross, ”data is just basically irrelevant,” and “Astro stuff was horrible”. Other students who rated theory low had less emotional responses, with Sean (T-2) and Wren (T-2) using words such as “a little disappointed,” “not fun,” “not entirely bad,” and “not horrible.”

On the other hand, students who rated computation as their low-interest method used a mix of positive and negative language. For instance, Bert (C-3) expressed that seeing programs run efficiently is really cool, but he doesn’t enjoy coding. He also admitted that he will need to use computation throughout his professional life, but he prefers to limit it as much as possible. Jamie (C-5) thought, “There’s less funkiness in computing, and slightly more funkiness in experiments,” while Moss (C-5) expressed having a love-hate relationship with it since “it has been frustrating working with computational physics...but useful.”

FIG. 2. Data analysis steps. (C) means a computation.
B. Learning Experiences

Within our data, students who were not interested in theory often had negative experiences with theory-related courses, both prior to and during college. Wren’s (T-2) low interest in theory was linked to their perception of high school geometry. They said, “Geometry class in high school was a lot of proofs, and I just did all the geometry proofs, and I was like, I don’t like doing proofs.” Additionally, one student shared his negative research experience as a contributing factor to his low interest. Armond (T-1) had a summer research experience in Astrophysics, which influenced his negative perception of theory. He said, “I didn’t really enjoy it. I really liked Astro before that. I was like ‘Astro is so cool, I’m going into Astro.’ But after I did that project, I completely got turned off by Astro. I was like, ‘I don’t want to do this ever again.’” The data is just basically irrelevant to me. It has no significance… It’s just numbers.” For Armond, a dislike of theory is coupled with a dislike of astrophysics. We will come back to the influence of irrelevance in the section on outcome expectations.

However, we found that students who rated computer lower often had negative experiences with coding during their research experiences. A few students also shared stories about their experiences with computational courses at various levels of education, from pre-college to BS/MS programs. It is important to note that not all of these learning experiences were necessarily negative or unpleasant. As we delve deeper into the results, we will examine what specifically made these experiences unpleasant in subsequent subsections.

C. Self-efficacy

When students were asked about their confidence in a method specialization that didn’t interest them, they often reported feeling capable to some extent. For example, Armond (T-1 & C-3) and Sean (T-2) felt confident in their understanding of physics. Negative perceptions of theory work, as seen in Armond’s case, could be linked to his negative experiences with computation, but he thought, “If I’m given the right tools, I could probably do it. I might be bored.” Kennedy (T-4) and Jamie (C-5), both rising 1st-year students, shared a similar sentiment that they were to some extent confident but also acknowledged that they still had more to learn. For example, Kennedy said, “I think we’ll get more confident over time as I understand more physics laws and stuff like that and be able to apply them to new solutions.” Dave (C-5), who became a physics major because he wanted a better understanding of the world, said, “If I actually tried, it could probably be pretty achievable. I would just learn how to code and then learn how to make predictive models.” However, for Bert, the mathematical process and coding were a particular challenge, and his confidence level varied depending on the complexity of the process. He said, “I had no clue how to go about coding, how to learn about it, how to do it, how it could improve, how I code, or what I could see that was wrong… Very minimally. I don’t understand them. In every other STEM category,…I can at least see how I did something wrong, whereas, for coding, there are no indicators of how your code is wrong. There is, but very minimally. I don’t understand them. My impression of coding was I couldn’t do this without a teacher, or I didn’t have enough motivation to do this without a teacher. Because I tried learning coding on my own at one point, and it just didn’t work. Either I just chose the wrong classes or just didn’t understand.”

A student’s self-efficacy in a method specialization can be influenced by pedagogy. For example, Moss (C-5) mentioned feeling more confident in one-on-one teaching situations. However, even in those settings, professors could negatively influence confidence if the student perceives the professor is down-talking or lecturing them.

D. Outcome expectations

Our research identifies three categories of outcome expectations that influence students’ interest development: disciplinary ideas, practices, and professional/personal life.

**Disciplinary ideas** refer to the in-depth content knowledge of a certain type of method specialization in physics. For example, Armond (T-1) who prefers “hands-on work rather than use his brain”, had difficulty grasping the micro-scale of the theory. Some students, such as Moss (C-5), had less interest in pursuing theory or computation because they were interested in the immediate translation to real-world applications rather than the acquisition of fundamental skills and knowledge. Sometimes in computation and theory, students may struggle with visualizing their work and not feel a physical connection to it. Sean (T-2) described it as “It can feel like just stringing ideas together” without anything concrete to hold on to. However, there are benefits to using computation in place of physical experimentation. For instance, it can save time and money. Armond (C-3) argued that by allowing the “computer to do the heavy lifting”, you just have to give it instructions. This can be challenging, but it can also be more cost-effective than conducting experiments over time.

**Practices** refer to the outcome expectations subcategory dealing with the day-to-day performance of particular skills or tasks associated with a method specialization. It is common for students to think that both computation and theory are math-intensive methods. Wren (T-2) feels that “theory involves a lot of proof” while Sean (T-2) finds it “unsatisfying to sit and do derivations.” Additionally, some students believe that these types of work won’t involve collaboration, as Dave (C-5) prefers “interacting with people and seeing the progress of hard work.” There is also a negative expectation that theory and computation work won’t be hands-on and will involve lots of sitting in front of the computer and lots of coding and formalizing ideas. For example, Wren (T-2) likes having their own thoughts but dislikes having to mathematically formalize them. Additionally, some students perceive working in theory to be “time-consuming,” with “lots of Zoom calls,” “long hours [of] reading,” “lots of literature searching and lots of generative writing,” “understanding textbooks,” “us-
ing a lot of chalkboards,” and “justifying each [idea].” For instance, Kennedy avoids theory because she “likes to figure things out quickly.”

Professional and personal life is an outcome expectations subcategory that addresses the overall lifestyle associated with a method rather than perceptions of the day-to-day work. As students think about their future careers, they consider their lifestyle expectations and how they can find a balance between their professional work and personal life. Many want to prioritize their mental well-being and create a healthy work-life balance. Kennedy (T-4), for example, hopes to have time for “social life and my other hobbies” while avoiding the stress of constantly thinking about work. While Moss (C-5) had a love-hate relationship with computation (i.e., frustrating but useful), they thought computational work would support their desired lifestyle, which is working remotely or traveling for conferences. They said, “When you’re a researcher, that’s essentially what I want to be, and when [you are] doing research, hopefully it would require me to travel sometimes.”

Theoretical work was usually linked to being a professor, working with students, and having an office in a physics department.

E. Proximal environment influences

Some students’ interests in computation may be influenced positively or negatively by environmental factors such as family or friends in the field. Moss (C-5), for example, was initially pressured by their father to pursue computer science, but the pressure actually deterred them from it as a teenager. However, as students’ interests can evolve over time, encouragement and support from loved ones can have a significant impact in reshaping their future career paths. In addition to being a role model for their little brother at home, Moss explained how they work in a nursing home and have opportunities to explain their research to “old folks” who are “completely impressed and proud”. Besides family and friends, Armond (C-3) and Dave (C-5) found it challenging to interact with other students during the COVID-19 pandemic, which lowered their interest in courses during that time.

F. Sense of belonging

Sense of belonging is an important element that is not part of the SCCT framework and focuses on the social connections of students within their community. Creating a sense of connectedness and being recognized by others through different learning experiences can enhance interest in a particular method. Moss (C-5) shared their experience of feeling out of place in the computational courses, where they rarely felt encouraged or congratulated. Moss also noticed an unequal representation of genders in computation classes, which made them feel like they didn’t belong in those spaces. However, they found more support in their research setting, where their peers and mentor provided them with more moral support, leading to a more positive experience with computational physics. Moss said once they did work, their peers were “like, Oh, my God, congrats, good job! They all send a little reaction, like a cute congratulations emoji, and that kind of stuff is important.”

On the other hand, lack of encouragement or negative experiences led to a lack of sense of belonging, as seen in Sean’s case, where he felt excluded in theory classes because his peers didn’t take him seriously. In addition, Sean (T-2) also stated that he had “to prove myself to the professors by the way that I look.” Sean did not elaborate on this statement.

V. CONCLUSION

The findings on students’ perceptions of theory and computation have important implications. The origins of students’ negative views can highlight specific areas where the curriculum could be enhanced. Despite frustration with coding and debugging, students with lower interest still recognized the importance of computational work in their future careers, emphasizing the practicality of integrating computational skills into the physics curriculum. On the other hand, students’ perceptions of theory as irrelevant and not useful led to negative outcome expectations. To address this gap, theory courses could incorporate more real-world applications and reflect the impactful areas of contemporary physics research. However, additional computational training may be needed to help students become more efficient at coding and debugging, which may make the experience less frustrating. Students perceived broad career options involving computational work, but that theoretical work was associated with being a professor. This may be partly correct and suggests that departments need to address the relevance of theoretical coursework to non-academic careers. Additionally, students need opportunities for recognition to build a sense of belonging with respect to different aspects of physics. Courses involving collaboration, regular formative feedback with encouragement, and projects may be more likely to support a sense of belonging.

By conducting further research, we hope to identify areas where students require additional resources. Our long-term goal is to create assessment tools that gather broader data, which can be used by the physics departments to identify sources of negative interests and whether some students lack access to positive experiences. Using this information, departments could design targeted strategies to improve support for students in their academic and professional journeys. Ultimately, we emphasize the importance of aligning curriculum, career guidance, and instructional approaches to enhance students’ interest and readiness for a wide range of careers in physics.

VI. ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under Grant No. 1846321.
Analyzing AI and student responses through the lens of sensemaking and mechanistic reasoning

Dean Zollman, Amogh Sirnoorkar, and James T. Laverty

Department of Physics, Kansas State University, Manhattan, Kansas - 66502

Physics education research (PER) shares a rich tradition of designing learning environments that promote valued epistemic practices such as sensemaking and mechanistic reasoning. Recent technological advancements, particularly artificial intelligence has caught significant traction in the PER community due to its human-like, sophisticated responses to physics tasks. In this study, we contribute to the ongoing efforts by comparing AI (ChatGPT) and student responses to a physics task through the cognitive frameworks of sensemaking and mechanistic reasoning. Findings highlight that by virtue of its training data set, ChatGPT’s response provide evidence of mechanistic reasoning and mimics the vocabulary of experts in its responses. On the other hand, half of students’ responses evidenced sensemaking and reflected an effective amalgamation of diagram-based and mathematical reasoning, showcasing a comprehensive problem-solving approach. In other words, while AI responses reflected how physics is talked about, a part of students’ responses reflected how physics is practiced during problem solving. We discuss the implications of this finding with an emphasis on epistemology of AI responses and designing next-generation assessments in physics.
I. INTRODUCTION

Physics education research shares a rich history of developing learning materials that promote valued epistemic practices. Sensemaking, the process of addressing perceived gaps in understanding, is one such practice [1]. Sensemaking represents one of the primary processes through which disciplinary experts generate new knowledge [2–4]. Sensemaking assists students in better content understanding [5] and facilitates in developing “sophisticated epistemology” [6]. Given this significance, there has been an uptick in investigations concerning the nature of discourse [7] and task features [8–10] that promote sensemaking in physics.

Artificial intelligence (AI) interfaces such as ChatGPT [11] have also gained increased traction in PER. These interfaces are software applications designed to generate human-like responses through Natural Language Processing. The “chatbots” interact with users by generating probabilistic responses to prompts based on large training data sets including physics texts [12]. Researchers have observed these interfaces to “pass” physics standardized assessments such as the Force Concept Inventory [13–15]. Noting this grammatical and technical sophistication, studies have called for exploring the ways in which AI can be leveraged in promoting valued epistemic practices in learning environments [16].

We contribute to these efforts by shifting the focus from analyzing conceptual correctness of AI responses to analyzing them within the frameworks of cognitive processes which are fundamental to scientific inquiry [1,17]. A better understanding of the AI responses in light of cognitive features can guide the effective integration of emerging technologies in our classrooms and laboratories. We compare the characteristics of responses generated by ChatGPT and introductory students to a physics problem specifically designed to foster the integration of physics principles with everyday ideas. We analyze the responses through the theoretical perspectives of sensemaking and mechanistic reasoning, along with noting the use of representations in making the requisite conclusions.

Findings highlight that by virtue of its training data, ChatGPT mimics the vocabulary of experts in its responses. On the other hand, students’ responses reflect an amalgamation of diagram-based and mathematical reasoning, showcasing a comprehensive problem-solving approach. We discuss the implications of these observations for future explorations especially along exploring the epistemology of AI responses and designing assessments. In doing so, we address the following research question: How does ChatGPT’s responses to a physics task compare to students’ responses in terms of the characteristics of sensemaking and mechanistic reasoning?

In the next section, we present the theory of sensemaking and mechanistic reasoning. Thereafter, in Section III, we discuss the physics problem, along with our methods of data collection and analysis. We then present the results of the relative comparison between students’ and AI responses on various cognitive features in Section IV before concluding with implications and future work in Section V.

II. THEORY

A. Sensemaking

In the rest of this paper, we adopt the following account of sensemaking from Odden and Russ [1]:

“a dynamic process of building or revising an explanation in order to ‘figure something out’ - to ascertain the mechanism underlying a phenomenon in order to resolve a gap or inconsistency in one’s understanding. One builds this explanation out of a mix of everyday knowledge and formal knowledge by iteratively proposing and connecting up different ideas on the subject. One also simultaneously checks that those connections and ideas are coherent, both with one another and with other ideas in one’s knowledge system.”

Based on the above definition, we identify the following markers (henceforth referred to as “sensemaking elements”) which together evidence the sensemaking process.

1. Noticing of inconsistency in understanding.
2. Blending everyday and formal knowledge.
3. Generating and connecting diverse ideas (e.g., conceptual, procedural, and intuitive).
4. Seeking coherence between the generated ideas.
5. Unpacking the mechanism of the phenomenon.

We analyze AI and students’ responses to our physics problem (Section III) through the lens of the above-mentioned elements. The fifth sensemaking element – unpacking the underlying mechanism of a phenomenon – however, is a complex process in itself. In the next sub-section we briefly detail what this process is, along with describing its markers.

B. Mechanistic reasoning

The process of unpacking the underlying mechanism of a phenomenon (also known as “mechanistic reasoning”) is a form of causal reasoning that entails description of the events and factors responsible for the occurrence of the phenomenon. Mechanistic reasoning entails generating explanations by moving from the observable features of the phenomenon to the underlying entities or processes, often at a micro level. The process of ascertaining the mechanism may also involve transitioning back from the micro to the macro-level features, as well as testing the validity of the explanation by altering the spatial or temporal organization of the involved entities or processes.
You are asked to design a Gravitron for the county fair, an amusement park ride where the rider enters a hollow cylinder, radius of 4.6 m, the rider leans against the wall and the room spins until it reaches angular velocity, at which point the floor lowers. The coefficient of static friction is 0.2. You need this ride to sustain mass between 25-160 kg to be able to ride safely and not slide off the wall. If the minimum \( \omega \) is 3 rad/s, will anyone slide down and off the wall at these masses? Explain your reasoning using diagrams, equations and words.

FIG. 1. Statement of the Gravitron problem

As an example, consider the phenomenon of the apparent shift in the position of a coin placed in a glass of water as viewed from above. A mechanistic account underlying this observation can be provided using the principle of refraction of light. When a ray of light passes from a denser medium (water) to a rarer medium (air), it deviates away from the normal drawn at the interface between the two media. This deviation results in the formation of an apparent image of the coin, making it appear closer to the water’s surface.

Krist et al. [18] describe the process of generating mechanistic explanations in terms of the three patterned strategic ways of knowledge-building called “epistemic heuristics”:

1. Thinking across scalar levels. Describing the actors identified and characterized at the scalar level below the observed phenomenon. In our example, transitioning to the level of light rays marks this heuristic.

2. Identifying and unpacking relevant factors. The description of the activities engaged by the identified actors at the lower scalar level. This can correspond to the bending of light rays relative to the normal and the formation of the apparent image of the coin.

3. Linking to coordinate relationships over time and space. Validating the generated explanation by varying the temporal and spatial organization of the involved actors and their activities. In our example, reversing the media and concluding that an object in the air medium viewed from the water would be perceived as farther away than its actual position.

We analyze mechanistic reasoning in students’ and AI’s responses through the lens of the above-mentioned heuristics.

III. METHODS

The objective of the current study is to compare students’ and ChatGPT’s responses to a physics problem which we refer to as the “Gravitron task” (Refer to Fig 1 for the problem statement). The task was designed using the Three-Dimensional Learning Assessment Protocol (3D-LAP) [19] to elicit the scientific practice of “Developing and Using Models” [20]. It involves a rotating cylindrical amusement ride in which a rider leans against the wall. Given the various parameters of the ride, the task asks to make a claim/prediction whether the rider would slide off the walls. Due to space constraints, we refrain from providing a detailed solution of the task and refer readers to our earlier work [21].

The students’ responses to the Gravitron task are derived from think-aloud interviews conducted in Spring 2018. The interviews involved ten introductory students individually solving a total of ten physics problems (with the Gravitation task being seventh on the list) while simultaneously articulating their thoughts aloud. The interview protocol involved asking the participants to treat the problem-solving exercise as an untimed exam. During moments of prolonged silence, the interviewer interjected with questions such as “What are you thinking?” to make students articulate their thoughts out loud. Of the ten, two students’ responses contained audio/video issues and thus are not part of this study. On the other hand, we recorded ChatGPT’s responses to the Gravitron task by using the same problem statement provided to students in the chatbot’s interface. In order to maintain symmetry in terms of the number of student responses, ChatGPT was prompted eight separate times by initiating distinct chat sessions, and the responses were recorded accordingly.

As part of the analysis, the students’ responses to the Gravitron task were first transcribed by taking into account their verbal arguments and written solutions. The second author then analyzed the transcripts through the lens of the five sensemaking elements described in Section II. The first element, noticing inconsistencies, was identified in the responses by noting puzzling questions [7], intermittent pauses, incomplete arguments [21, 22], etc. The second sensemaking element, blending everyday and formal knowledge, was identified when the responses reflected amalgamation of formal physics principles with the Gravitron’s physical system. Arguments such as “the rider being pulled down by the gravity” or “friction preventing the rider from slipping down” were coded as the second element. The third and fourth elements were captured based on the generation and validation of intuitive, conceptual, or procedural ideas.

We captured the fifth sensemaking element (mechanistic reasoning) by identifying the three epistemic heuristics described in Section II. Students’ arguments concerning forces (and other quantities) were coded as the first epistemic heuristic. However, students who engaged only in plug-and-chug of equations, even though the equations entailed force terms were not coded for this heuristic. Identifying the relevant forces (and other quantities) and their interaction on the Gravitron’s riders were coded as the second heuristic of mechanistic reasoning. Lastly, transitioning to the macroscopic Gravitron scenario to make the requisite claim marked the final heuristic in the students’ and chatbot’s responses. In addition, since the problem statement explicitly asked for relevant representations used in making the claims, and the contemporary literature notes the role of representations in mechanistic reasoning [23], we also note diagrams and equations employed.
in the responses. In summary, the analysis of students’ and ChatGPT’s responses involved (i) identifying the first four sensemaking elements, (ii) the three epistemic heuristics of mechanistic reasoning (the fifth sensemaking element), and (iii) representations such as diagrams and equations.

We wish to highlight a few potential limitations in our methodology. First, the nature of responses across the two sets are inherently not identical. And two, since the authors were aware of the source of the solutions, we acknowledge there could be a likely bias while analyzing the responses.

IV. RESULTS

We now discuss how the sensemaking elements manifest in students’ and AI’s responses to the Gravitron task. These results have also been summarized in Table I.

1. Noticing inconsistency in understanding

The meta-cognitive activity of noticing gaps in one’s knowledge system forms a central feature of the sensemaking process. Four of the eight students’ responses reflected evidence of noticing this gap (two of which have been detailed in References [21, 22]). None of the ChatGPT responses reflected this sensemaking element. We believe the implication of this result needs to be reemphasized as it is easy to implicitly model ChatGPT as a “person responding from the other end”. As Kortemeyer [13] notes

“It is irritatingly hard not to anthropomorphize ChatGPT. As a physics teacher, one invariably finds oneself rooting for the students and thus by extension also for ChatGPT, celebrating its successes and being frustrated about its occasionally inexplicable failures.”

By virtue of its design, the chatbot generates probabilistic responses from its training data set and at least currently does not possess the meta-cognitive ability to address gaps in its “knowledge system”.

2. Use of everyday and formal knowledge

As a task based on real-world scenario, a key feature of the Gravitron task is its potential to facilitate the amalgamation of the physics principles with everyday ideas. Six of the eight students’ responses entailed arguments which reflected how various forces (and other quantities) interacted in holding up the Gravitron’s rider. On the other end, all of the eight ChatGPT’s responses elicited this mode of reasoning (albeit some were conceptually incorrect).

3. Generating and connecting ideas

The Gravitron task requires leveraging conceptual and procedural ideas from both physics and mathematics. Five of the students’ responses and all eight of the chatbot’s responses reflected this feature. Both sets of responses included arguments involving diverse sets of physical quantities such as forces, accelerations, and momenta along with their directionality. However, students’ responses differed from ChatGPT in two ways. They reflected intuitive arguments without any conceptual or procedural justification. Students’ responses also included incomplete arguments which were either changed or terminated abruptly. Such iterative construction and/or revision of arguments has been noted as a characteristic feature of students’ knowledge building process [1].

4. Seeking coherence between ideas

The generated ideas were then brought together in making the required claim about the riders’ status inside the Gravitron. Half of the students and all eight responses from ChatGPT connected and/or justified the solution back in terms of the earlier invoked ideas. However, we observed two key differences between the students’ and AI’s concluding statements. Interestingly, all of the ChatGPT’s responses made the incorrect conclusion that riders would not fall off the Gravitron’s walls (as compared to four of the students’ conclusions). This observation assumes significance as the chatbot does not “solve” the given problem, but rather produces the most probable result from its training data set. This result indicates that despite all its sophistication in the previous two sensemaking elements, the conclusions from AI should not be taken at its face value. Secondly, three of the ChatGPT conclusions accompanied detailed assumptions under which the conclusions hold true.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Students (Out of 8)</th>
<th>ChatGPT (Out of 8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensemaking</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>Mechanistic reasoning</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Noticing gaps in understanding</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>Blending everyday and formal knowledge</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Generating and connecting ideas</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Seeking coherence between the ideas</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Epistemic heuristics of mechanistic reasoning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thinking across scalar levels</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Identifying and unpacking relevant factors</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Linking to coordinate relationships over time and space</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Representations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagrams</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Equations</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>
5. Generating mechanistic explanations

5.1 Epistemic heuristics

We see symmetry in both sets of responses on this aspect. While only half of the students’ solutions reflected the three epistemic heuristics, all of ChatGPT’s responses reflected the “mechanistic sophistication” [17].

5.2 Use of representations

One of the contrasting differences between the two data sets is in terms of diagram-based reasoning. While seven of the eight students’ written responses entailed diagrams, only three were noted in ChatGPT’s responses. Figure 2 represents one of the three diagrams generated by ChatGPT. The AI-generated diagrams differed from student ones on two features. The AI representations (unlike the student-generated diagrams) did not represent the rider-Gravitron system by highlighting the interaction of relevant forces. And secondly, the responses did not leverage diagram-based arguments into their conclusions. Rather all the three diagrams were generated after making the requisite conclusion. Along with diagrams, equations were the other form of the focused representations. On this end, all eight responses from students and ChatGPT contained equations.

V. DISCUSSION AND CONCLUSION

We analyzed students’ and ChatGPT’s responses to the Gravitron task (Figure 1) in terms of (i) sensemaking, (ii) mechanistic reasoning, and (iii) the use of diagrams/equations. The findings (summarized in Table I) highlight two key contrasting features between the two sets of responses. Most of the students’ reasoning highlight richness in the practices of the problem-solving exercise, especially in blending diagrams with mathematical arguments. That is, students’ responses reflected the practices involved in engaging with physics. On the other hand, since ChatGPT’s responses are drawn from its training data set, they mimic the lexical patterns or vocabulary of an expert’s reasoning in physics (despite providing incorrect answers in all of its solutions). In summary, while students’ responses reflect how physics is practiced, the AI’s responses reflect sophistication in how physics is talked about during problem solving.

Of the eight students, four evidenced sensemaking and mechanistic reasoning in their responses. On the other hand, all eight AI’s responses evidenced the elements of sensemaking (including mechanistic reasoning) with an exception of the meta-cognitive feature of noticing gaps in understanding. This should not come as a surprise as the AI interface is not capable (at least at the moment of writing this paper) of engaging in any meta-cognitive activity. Despite the strong evidence in characteristics of sensemaking and mechanistic reasoning, none of the AI’s solutions provided a correct prediction or conclusion. This observation calls for attention on two aspects. One, by virtue of its training data, the AI mimics the vocabulary of an expert on how physics is talked about during problem solving. Two, an incorrect answer put forth through sophisticated argument calls for diligence against conflating the conceptual merit of AI’s argument with its semantic sophistication. This claim is also in agreement with contemporary studies calling for careful evaluation of AI responses, particularly in terms of scientific accuracy [24].

Yet another contrasting feature between the two sets of responses is difference in the employment of representations during problem solving. While seven of the eight students employed diagrams, only three of the ChatGPT’s responses reflected the same. Furthermore, unlike the student-generated diagrams, the AI diagrams did not portray the appropriate Gravitron-rider system with relevant forces and were not integrated in the solutions. Thus students’ responses to the Gravitron task reflected the practice of blending diagram-based reasoning with mathematical arguments thereby reflecting the key practices of engaging with physics.

For instructors, our observations provide insights on designing assessments using AI. The assessments can focus on asking students to critique and validate AI-generated solutions. Such exercises would serve in promoting the sophisticated vocabulary of talking about physics while simultaneously guarding students against conflating the lexical sophistication with conceptual correctness. For researchers, our observations call for investigating the epistemological messages [25] about learning physics conveyed by AI responses to students. That is, unpacking the implicit messages conveyed by AI to students on what counts as “knowing” or “doing” science. Exploring AI responses to well-structured and ill-structured problems also present a potential avenue.

Our future work would involve extending similar methodology in analyzing students and AI solutions across other epistemic practices such as modeling and argumentation. Given the evidence of epistemic heuristics of mechanistic reasoning in AI’s responses, we also seek to explore the solutions in finer details using alternate frameworks [17] and across enhanced version of ChatGPT and other AI interfaces.

VI. ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 2013339.


PERC 2023 Presentations

Community Connection Abstracts

Coble, Kim, Intro Astronomy Connections
Demaree, Dedra, Alternative Careers for PER Graduates: High School Teachers
Henderson, Charles, Meet with your Journal Editor I: Phys Rev PER
Lindell, Rebecca, Disabled Physicists Meetup
McDermott, Liam, LGBTQIA+ PER Meetup
McDermott, Liam, New/Unconnected Researcher Connection, Collaboration, and Creation Space
Morphew, Jason, Making Connections in Cross-Discipline Based Education Research (XDBER)
Scanlon, Erin, Alternative Careers for PER Graduates: Primary Teaching Positions at Large R1 Universities
Urquhart, Mary, Connecting PER to K–12
White, Gary, Meet with your Journal Editor II: The Physics Teacher
Zich, Raymond, Solo PER Community Connection

Contributed Poster Presentation Abstracts

Abdurrahman, Fatima, A case study of tensions in student-faculty partnerships for departmental change work
Alfson, Jonathan, Analyzing the functions of multiple external representations of electric potential
Allen, Josephine R., Understanding students’ struggles with collaboration through their views of knowing
Allen, Winter Rose, Using natural language processing to predict correctness of students’ problem-solving strategy essays
Arnell, Jared, Air particles in a lattice: Considerations for sound wave simulations in physics education
Aryal, Bijaya, Exploring Students’ Learning Expectations in a Two-phased Instructional Model
Ayouz, Mehdi Adrien, Teaching quantum mechanics in an experiential learning and actively engaging environment
Bagdovitz, Peter, Modeling confusion in collaborative learning
Bano, Roshni, “Science happens between people”: teachers’ perspectives in a physics RET program
Barth-Cohen, Lauren, Acquisition of qualitative video data: methods and reflections in PER
Bauman, Lauren, Shifts in students’ responses to conceptual questions after a new physics conceptual worksheet: Preliminary findings
Borish, Victoria, Student perspectives about seeing quantum effects in experiments
Bralin, Amir, Analysis of student essays in an introductory physics course using natural language processing
Bridges, Bill, Investigating the Assessment Landscape of Physics Graduate Programs
Brundage, Mary Jane, Investigating context dependence of introductory and advanced student responses to introductory thermodynamics conceptual problems
Buncher, John B., Effects of Anchor Item Choices on Bias on the Force Concept Inventory across the intersection of gender and race
Burkholder, Eric, Defiance in the face of adversity: a qualitative study of women's attrition from and persistence in physics
Burns, Andrew, Emergent Explicit Regulation in Collaborative College Science Classrooms: A Case Study with ASL
Cano, Ana Gaby, Neurodivergent student experiences in ungraded classes
Cantlin, Kevin, Exploring the impact of low-stakes assessment in introductory physics
Cao, Ying, Pre-service Teachers’ Understandings of Scientific Inquiry
Chatta Subramaniam, Ravishankar, Characterizing the 'design-science gap' in an engineering design-based laboratory unit in an introductory physics course for future engineers
Chaudhury, S. Raj, *Actors, Scripts and Orchestration: a framework for understanding Team Based Inquiry Learning*

Corpuz, Edgar, *Physics students’ perceptions of team-based learning*

Dalka, Robert P., *Considering the Departmental Action Leadership Institute as a Community of Transformation: What’s highlighted and what’s missed?*

Dana, L., *More than technical support: Professional contexts of physics instructional labs*

Dancy, Melissa, *Physics instructors have limited ability to recognize or address bias in student-student interactions*

Doty, Constance, *Using clusters of models of disabilities to describe support for mentees with disabilities*

Drury, Byron, *Building a mentorship community of practice*

Emigh, Paul J., *Student belonging in STEM courses that use group work*

Engblom, Samuel W., *Learning assistants’ teaching strategies for promoting scientific inquiry among undergraduate students in a physics laboratory setting*

Erukhimova, Tatiana, *Broadening Student Learning through Informal Physics Programs*

Franklin, Maxwell, *Correlating Attitudes with Persistence in Undergraduate Women*

Fredly, Karl Henrik, *Development of Computational Literacy in a Computational Physics Master’s Program*

Freeman, M. T., *Applying Voting Theory to Mastery Grading; A Study of Faculty Interpretation of Course-Level Categorical-Score Distributions*

Freeman, Niah, *The Impact of Reflective Journaling on STEM Undergraduates in Physics & Astronomy Labs*

Garcia, Tyler, *Interview validation of tasks assessing knowledge-in-use*

Geschwind, Gayle, *Representational differences in how students compare measurements*

Goldak, John, *Effect of Essential Skills Practice on Student Understanding of Probabilities in an Upper-Division Quantum Mechanics Course*

Griston, Molly, *A characterization of existing quantum mechanics assessments*

Guo, Qirui, *The Journey of Quantitative Literacy Development: Insights from Physics Majors*

Guthrie, Matthew W., *Investigating experiences of marginalized physics students in group work settings*

Hackler, Stephen, *Development and Reliability Analysis of a Split-Administration Test of the Mathematical Epistemic Games Survey*

Hamed, Razan, *Relating visual attention and learning in an online instructional physics module*

Henchy, Colin, *Reducing belief in pseudoscience and conspiracy theory among undergraduate students with PACEM (Pseudoscience And Conspiracy-theory Educational Module)*

Henriquez, William, *Examining socio-metacognitive skills of students while struggling with experiment setting*

Herman, Jyllian, *Investigating the role of question-asking and making contributions in bridging cultures to foster physics identity development*

Horak, Maria Sue, *A Look Physics Teacher Identity Around Equitable Instruction: The Tour Guide, Coach, and Gardener*

Hu, Peter, *Development, validation and online and in-person implementation of clicker question sequence on change of basis*

Hughes, Mark, *Do Students Think an Object Can Ever Have a Definite Position?*

Hull, Michael M., *Simulated Groupwork in an Asynchronous Course Learning about Radioactivity*

Huynh, Tra, *Drawing on force ideas for kinematic reasoning in introductory physics*

Ismael, Safana, *Toward helping students develop error detection skills*

Juanson, Adrian, *Research on a faculty support program for working with learning assistants*

Kramer, Hannah M., *Understanding student difficulties with VPython in University Physics through framing*

Kruse, Jesse, *Rasch Analysis of the Quantum Mechanics Concept Assessment*

Kustina, Jonathan, *Metacognitive knowledge and regulation of peer coaches*

Lakshmanan, Aakash, *Effects of Pre-lab Video on Z-score Understanding in Introductory Labs*

Le, Thanh, *Exploring alternative perspectives through fictionalized student dialogues*
Leukes, Chad, Probing student sense-making in the introductory astronomy classroom
Li, Yangqiuting, The effects of active learning on students’ sense of belonging and academic performance
Lindell, Rebecca, Fluids Conceptual Evaluation Update: Sign up for Pilot Test!
Ma, Guofu, Concept Inventory for Upper-level Undergraduate Students in Lagrangian and Hamiltonian Mechanics
Machrone, Alexandra, Assessing Sense of Belonging, Climate, and Culture at the Departmental Level from the Perspective of Undergraduate Students
Maldonado, Danielle, Adding self-regulated learning instruction to an introductory physics class
Maries, Alexandru, Upper-level students’ conceptual understanding of energy and momentum
Mashood, K. K., Mapping the landscape of physics education research in India: Preliminary insights
Mathis, Clauseill, Exploring sense of belonging and imposter phenomenon among physics teachers
May, Jason, Multimodality and Video Research in PER: A Methodological Case Study in Intro Physics Labs
McColgan, Michele W., Augmented Reality to Scaffold 2D Representations of 3D Models in Magnetism
McDermott, Liam G. E., “Academia as a whole is structured entirely without any consideration for neurodivergency,” and other things neurodivergent students want you to know
McHale, Sarah, Designing a Survey to Characterize Student Perceptions of Computation
McKagan, Sam, PhysPort Redesign
McPadden, Daryl, Planning for Participants’ Varying Needs and Abilities in Qualitative Research
Megowan-Romanowicz, Colleen, Evaluating learning of motion graphs with a LiDAR-based smartphone application
Meredith, Dawn, Injecting Fluids into Your Introductory Physics Course for Life Science Students
Merzel, Avraham, Diagnosing High School Students’ In-lesson Difficulties in Quantum Physics
Meyer, Josephine C., Identifying assessable content for Quantum Computing Conceptual Survey
Mirabelli, Joseph F., Stressors for First-Year Doctoral Physics Students: A Qualitative Case Study
Monsalve, Camila, A Curander@s journey to support two-year college STEM students’ in finding their scientific belonging
Montgomery, Barron, How traditional physics coursework limits problem-solving opportunities
Moore, Tony, I’m not an Ableist! Am I?
Mueller, Sara, Labs are fun: redesigning the advanced instructional lab to bring joy to students and instructors
Myers, Carissa, Leveraging Daily Journal Prompts to Capture the Complexities of Self-Efficacy: Coding Across Iterations
Nandivada, Urja, A discussion on attitude shifts in First Year Physics and our solution
Nandivada, Urja, PERCoG’s Communitree! An Interactive Poster
Nanthou, Delwrick, Grappling with the dominant narrative of physics: Teachers rethink colonial roots together to reshape classrooms
Ndihokubwayo, Kizito, Characterizing students' participation rates in LASSO before and after instructor motivational practices
O’Donnell, Christine, Supporting first- and second-order departmental change with the Effective Practices for Physics Programs (EP3) Guide
Oleynik, Daniel, Three axes for expressing disability models and experiences: The cause, the effect, and the ability/disability dichotomy
Oliver, Travis, Self-efficacy growth in a standards-based assessment and active learning classroom
Ouellette, Ellen, Alignment between student epistemological views and experiences with course structures in introductory physics: A case study
Owen, Jon, "Which has more energy?" - An example of responsive teaching in university physics
Pacheco, Daniel, Analyzing students’ sense of belonging and physics identity in a modern physics class
Parobek, Alexander (Alex) P., Students’ interpretations of disciplinary convention with the first law of thermodynamics
Patterson, Zac, Secondary students’ epistemic framing of contemporary physics
Pedicone, Michael, “Why waste time say lot word when few word do trick?”
Perwaiz, Fatima, Productive Failure Learning with a Contrasting-case Design in Undergraduate Physics
Piedrahita U, Yuri B., Exploring the scientific argumentation quality level of upper vs lower performance undergraduate students in an online physics laboratory
Poulos, Parker E, Instructor Expectations and Reactions Toward Assessment Feedback
Prakash, Harish Moni, A new method to determine rapid response time cutoffs for student data trimming
Provenzano, Nicolas, Creating Inclusive Environments Through Peer Mentoring
Quichocho, Xandria R., LA Program as a Driving Force for Identity Development Through Access to Ideational Resources
Reisman, Ezra, Comparing students’ ideas about collaboration with socio-metacognition
Robbins, Jocelyn, The Tentativeness and Trustworthiness of Science
Roth, Drake, Student affective experiences in introductory physics for life sciences
Ryder, Diana, Tools for Understanding the Microscopic World of Quantum Mechanics: Analogies in Textbooks
Santana, Lisabeth Marie, Investigation of students repeating introductory physics courses by ethnicity/race
Santoso, Purwoko Haryadi, Predicting physics students’ achievement using in-class assessment data: A comparison of two machine learning models
Scanlon, Erin, Describing a disability faculty online learning community: Instructors’ views about inclusive teaching strategies
Scheuneman, Stacy M., Characterizing representational gestures in collaborative sense-making of vectors in introductory physics
Sengul, Ozden, College Science Students’ Epistemology in Solving Physics Problems
Shafer, Devyn, Shifts in Student Support Structures from High School to College Engineering Programs: A Longitudinal, Qualitative Study
Sharkey, Daniel, Physics Graduate Teaching Assistant Use of Error Framing in Recitations and Laboratories
Sirnoorkar, Amogh, Analyzing students’ assumptions to varying degree of prompting during problem solving
Sivitilli, Alexander, Investigating how students engage with a digital planetarium
Smith, Mathilda Jacqueline, Unlearning indoctrination: Tensions between decolonizing curricula and characteristics of whiteness
Snow, Al K., A case in which canonically incorrect ideas do not hinder conceptual progress in introductory physics
Solorio, Christian D., Conceptual Challenges of Discretizing Wave Functions: A Case Study
Speirs, J. Caleb, Metacognitive Calibration and Student Self-evaluation on (Un)graded Quizzes
Stanley, Bryan, Informal physics volunteers: Motivations from student to career in education
Steyn, Mayhew, Challenges in designing remote physics education activities using students’ smartphones as measurement devices.
Stone, Antoinette, Student epistemologies in the post pandemic era
Sulaiman, Nidhal, Effect of online laboratory teaching on students views about the nature of science and physics experiments during COVID19
Sunil, Harshini, Investigating Academic Burnout in Undergraduate Physics Experiences
Sword, Astra, PER-QT: Towards a better PER publication pipeline by building a Quarto extension for PER journals
Sword, Astra, Testing for over- and under-dispersion in physics degree outcomes
Talafian, Hamideh, Barriers to Implementing Open-ended Labs with iOLab among High School Physics Teachers
Talafian, Hamideh, Characterizing the complexities of experimental decision making in an introductory lab practical
Tempkin, Jenna, Using Community Cultural Wealth to Understand Experiences in Physics Bridge Programs
Tjia, Angelina, Development of physics affinity in introductory physics for life science students across three institutions
Trojand, Daniel, Student decision making in the undergraduate lab program
Verostek, Mike, Inequities and misaligned expectations in PhD students' search for a research group
Vignal, Michael, Couplet Scoring: A New Paradigm for Designing and Scoring Research-Based Assessment Instruments
Wagner, D. J., Development of Reynolds-Number questions for the Fluids Conceptual Evaluation (FCE)
Waterson, Alyssa, Using Network Analysis to Represent Transfer Student Cohorts
Watkins, Jacob, Investigating Perceptions of Relevance Towards Computation in an Introductory Physics for Life Science Course
Wheeler, Sav, ChatGPT reflects student misconceptions in physics
Willison, Julia, Examining faculty choices while implementing the Next Gen PET curriculum through Revealed Causal Mapping
Wu, Xian, Analyzing the dimensionality of the Energy and Momentum Conceptual Survey using Rasch Measurement Theory
Young, Tamara Gay, Students' use of symmetry as a tool for sensemaking
Zich, Raymond, Extended evaluation of the instructional effectiveness of computational exercises in teaching general education astronomy courses
Zohrabi Alaee, Dina, Analyzing Physics Majors' Specialization Low Interest Using Social Cognitive Career Theory
Zollman, Dean, Analyzing AI and student responses through the lens of sensemaking and mechanistic reasoning

Critical Conversation Abstracts

Lui, Kristine, Doing PER in Partnership with Two-Year Colleges
McKagan, Sam, Comparing different models for instructional materials based on a resources framework
Stelzer, Tim, Referee Guidelines to maintain clarity and generativity of quantitative research studies in Physical Review Physics Education Research

Plenary Abstracts

Barbato, Lyle, PERC Proceedings Timeline
Cid, Ximena, The PER Community: The known and unknown
Coble, Kim, Empowering scholars to change the culture of STEM pedagogy to be more inclusive and equitable
Mestre, Jose P., Reforming Large STEM Introductory Courses Through Communities of Practice
Zollman, Dean, The birth and early growth of the Physics Education Research Conference

Round Table Discussion Abstracts

Dachille, Frank, Exploring Culture in Physics
Tasar, Fatih, The “International Handbook of Physics Education Research”: Insights, Themes, and Future Directions
Wagner, D. J., Research-based Assessments in the Digital Age: Validity, Reliability, Accessibility

Symposium Poster Abstracts

Cano, Ana Gaby, Neurodivergent student experiences in ungraded classes
Lane, W. Brian, Physics Teachers' Motivations to Learn Computational Thinking as a Re-novicing Experience
Lane, W. Brian, Student Perceptions of Grading Practices
Nguyen, Sarah, Examining student experience and agency on ungraded quizzes
Odden, Tor Ole B., Theoretical frameworks for understanding computation in physics education
Paul, Cassandra, Exam retakes and revisions: Impacts on student outcomes
Pollard, Benjamin, Preliminary results from ungrading in large introductory physics courses
Sabo, Hannah C., Gentle assessment: a comparison of assessment of physics and computational physics
Swartz, Mark, *Interpreting graded problem solutions: The inconsistent messages that students receive*

**Symposium Talk Abstracts**

Allen, Scott E., *Transformative Instructional Reform at the Naval Nuclear Power School*

Ansell, Katie, *Informing the PER Discourse about Groups with Educational, Sociological, and Psychological Frameworks*

Chasteen, Stephanie, *Common Markers of Success within Reform Campaigns*

Dew, Matthew, *The Impact of Heterogeneity on the Benefits of Group Work in Lab*

Ives, Joss, *Two different approaches to measure the impact of gender composition in the performance of self-selected groups during group exams*

Oliver, Kristin A., "*The prettiest photos are the ones that have happy people in them": the use of photovoice in an upper-division physics capstone project course*

Pollock, Stephen, *Systematic Reform at CU Boulder*

Stelzer, Tim, *Long-Lived Reform at University of Illinois*

Sundstrom, Meagan, *Quantifying student collaboration in labs using social network analysis*

**Workshop Abstracts**

Izadi, Dena, *Using MAXQDA Software Effectively and Efficiently for Qualitative and Mixed Method Data Analysis*

Lindell, Rebecca, *Using AI to Streamline your Physics Education Research Practice*

Wu, Xian, *Fostering Effective and Inclusive Group Work*
PERC 2023 Author Index

Abdurrahman, Fatima N.: 10, 76, 421
Adams, Adrian L.: 46
Alesandrini, Anne T.: 52, 151
Alfson, Jonathan W.: 16, 421
Allen, Josephine R.: 22, 421
Allen, Scott E.: 426
Allen, Winter Rose: 421
Allie, Saalih: 308
Alvarado, Carolina: 22
Ansell, Katherine: 362, 426
Arnell, Jared: 28, 421
Aryal, Bijaya: 421
Avena, Jennifer S.: 169
Ayouz, Mehdi Adrien: 421
Bagdovitz, Peter: 34, 421
Bano, Roshni: 40, 421
Barron, Joshua: 64
Barth-Cohen, Lauren A.: 46, 403, 421
Bauman, Lauren C.: 52, 151, 320, 421
Bonn, Doug: 163
Boone, Téa: 368
Borish, Victoria: 254, 421
Boudreaux, Andrew: 22, 193
Braden, Sarah K.: 403
Bralin, Amir: 58, 344, 421
Bridges, Bill: 64, 421
Brown, Michelle N.: 132, 236, 314
Brundage, Mary Jane: 70, 205, 421
Buncher, John B.: 421
Burkholder, Eric: 338, 421
Burns, Andrew: 421
Cano, Ana Gaby: 421, 425
Cantlin, Kevin: 421
Cao, Ying: 421
Casey, Jennifer R.: 88
Chakarov, Alexandra Gendreau: 169
Chasteen, Stephanie Viola: 242, 426
Chatta Subramaniam, Ravishankar: 344, 421
Chaudhury, S. Raj: 422
Chini, Jacquelyn: 64, 82, 248, 296, 391
Cid, Ximena: 425
Close, Eleanor: 278
Coble, Kim: 425, 425
Cochran, Geraldine L.: 368
Corbo, Joel: 76
Coppaz, Edgar: 422
Craig, David A.: 76
Daane, Abigail R.: 132, 236, 314
Dachille, Frank Dachille: 425
Dalka, Robert P.: 10, 76, 422
Dana, L: 422
Dancy, Melissa: 422
Day, James: 163
Demaree, Dedra: 425
Dew, Matthew: 426
Donaldson, Jonan Phillip: 108
Doty, Constance M.: 82, 248, 296, 422
Drury, Byron: 422
Emigh, Paul J.: 16, 88, 422
Engblom, Samuel W.: 95, 422
Erukhimova, Tatiana L.: 108, 422
Flood, Virginia: 290
Franklin, Maxwell: 422
Fredly, Karl Henrik: 422
Freeman, Michael T.: 101, 422
Freeman, Niah: 422
Galanti, Terrie M.: 187
Garcia, Tyler: 422
Garrett, Carlee: 108
Geschwind, Gayle: 114, 422
Gire, Elizabeth: 16, 326
Goldak, John: 120, 422
Goodhew, Lisa M.: 52, 266, 320
Griston, Molly: 422
Gumale, Ahmed A.: 236
Guo, Qirui: 422
Guthrie, Matthew W.: 397, 422
Hackler, Stephen: 422
Hamed, Razan: 126, 422
Hansen, Brynna: 266
Harrer, Benedikt: 290
Hassel, George E.: 211
Henchy, Colin: 422
Henderson, Charles: 425
Henderson, Rachel: 64
Henriquez, William: 22, 422
Herman, Jyllian: 422
Heron, Paula R. L.: 320
Hinko, Kathleen: 322, 380
Holzinger, Eva: 145
Horak, Maria Sue: 132, 422
Hu, Peter: 138, 422
Hughes, Mark: 422
Hull, Michael M: 145, 422
Huynh, Tra: 151, 422
Hyater-Adams, Simone: 278
Ismael, Safana: 157, 422
Ivanov, Nikolay: 380
Ives, Joss: 163, 426
Izadi, Dena: 426
Jeidler, Maximilian: 145
Johnson-Glenberg, Mina C.: 224
Juanson, Adrian: 169, 422
Kita, Katsuhiro: 88
Kraft, Aaron M.: 163
Kramer, Hannah M.: 422
Krishna, Sujata: 88
Krist, Christina: 260
Kruse, Jesse: 175, 422
Kryjevskaia, Mila: 157
Kuo, Eric: 260
Kustina, Jonathan: 181, 422
Lakshmanan, Anakash: 422
Lane, W. Brian: 187, 425, 425
Laverty, James: 64, 101, 302, 415
Le, Thanh: 422
Leukes, Chad: 423
Lewandowski, Heather J.: 114, 254
Lewsirirat, Sarat: 260
Li, Yangqutuq: 138, 423
Li, Yaoguang: 397
Liao, Jiehong: 88
Lindell, Rebecca: 211, 423, 425, 426
Lui, Kristine: 425
Lundsgaard, Morten: 260
Lê, Thanh: 22, 193
Ma, Guofu: 423
Machrone, Alexandra: 423
Mahmood, Maggie S.: 95
Maldonado, Danielle: 199, 423
Maries, Alexandru: 205, 423
Mashood, K K: 423
Mathis, Clausell: 132, 236, 314, 423
Matos, Michael J.: 95
May, Jason M.: 46, 423
McCoy, Michele W.: 211, 423
McHale, Sarah: 423
McKagan, Sam: 423, 425
McPadden, Daryl: 423
Megowan-Romanowicz, Colleen: 224, 423
Meltzer, David: 70
Meredith, Dawn: 181, 423
Merzel, Avraham: 423
Mestre, Jose P.: 425
Meyer, Josephine C.: 423
Miller, Casey: 374
Mirabelli, Joseph F: 423
Mohan, Kirtimaan: 380
Monsalve, Camila: 423
Montgomery, Barron J.: 230, 423
Moore, Tony: 423
Morphew, Jason: 58, 211, 344, 425
Mosley, Nazeer A.: 217
Mueller, Sara: 423
Munsell, Jeremy: 126
Myers, Carissa: 423
Nandivada, Urja: 423, 423
Nanthou, Delrick: 132, 236, 423
Ndihokubwayo, Kizito: 423
Nguyen, Sarah: 425
Nissen, Jayson: 34, 88, 193
O'Donnell, Christine: 242, 423
Odden, Tor Ole B.: 425
Oleynik, Dan P.: 82, 248, 423
Oliver, Kristin A.: 254, 426
Oliver, Travis: 423
Ouellette, Ellen: 260, 423
Owen, Jon C.: 266, 423
O'Brien, Daniel J.: 223
Pacheco, Daniel: 423
Parobek, Alexander (Alex) P.: 272, 423
Patterson, Zac: 424
Paul, Cassandra: 425
Pedicone, Michael: 424
Perry, Jonathan D.: 108
Perwaiz, Fatima: 424
Piedrahita U, Yuri B.: 424
Pollard, Benjamin: 425
Pollock, Stephen: 426
Poulos, Parker E: 424
Piedrahita U, Yuri B.: 424
Pollard, Benjamin: 425
Pollock, Stephen: 426
Poulos, Parker E: 424
Prakash, Harish Moni: 424
Price, Argenta M.: 230
Provenzano, Nicolas: 424
Quan, Gina: 169
Quichocho, Xandria R.: 278, 424
Rebello, Carina: 58, 344
Rebello, N. Sanjay: 58, 126, 344
Reisman, Ezra: 424
Robbins, Jocelyn: 424
Robertson, Amy: 52, 151, 320
Roth, Drake: 424
Rozas, XL: 187
Ryder, Diana: 284, 424
Ryu, Minjung: 40
Sabo, Hannah C.: 425
Sachmpazidi, Diana: 10, 76
Saitta, Erin K. H.: 296
Santana, Lisabeth Marie: 424
Santoso, Purwoko Haryadi: 424
Scanlon, Erin: 82, 248, 391, 397, 424, 425
Scherr, Rachel: 386
Scheuneman, Stacy M.: 290, 424
Sebastian, Ryan Biju: 260
Sengul, Ozden: 424
Shafer, Devyn: 424
Shaffer, Peter: 120
Sharkey, Daniel: 296, 424
Shepherd, Matthew: 338
Singh, Chandrakula: 70, 138, 205
Sirnoorkar, Amogh: 101, 302, 415, 424
Sivitilli, Alexander: 308, 424
Smith, Mathilda Jacqueline: 314, 424
Snow, Ai K.: 52, 320, 424
Solis, Caitlin: 64
Solorio, Christian D.: 326, 424
Sorby, Olin: 151
Speirs, J. Caleb: 424
Stagnitti, Natalie C.: 211
Stanley, Bryan: 332, 424
Stelzer, Tim: 425, 426
Stewart, John: 199
Steyn, Mayhew: 424
Stone, Antoinette: 424
Strain, Rebecca Smith: 338
Sulaiman, Nidhal: 424
Sundstrom, Meagan: 426
Sunil, Harshini: 350, 424
Swanson, Hillary: 28
Swartz, Mark: 426
Sword, Astra: 356, 424, 424
Talafian, Hamideh: 362, 424, 424
Tasar, Fatih: 425
Tempkin, Jenna P.: 368, 424
Tjia, Angelina: 424
Tonry, Keegan Shea: 409
Towns, Marc H.: 272
Trojand, Daniel: 425
Turpen, Chandra: 10, 76
Urquhart, Mary: 425
Vargas, Jackie: 95
Verostek, Michael: 284, 374, 425
Vieyra Cortés, Christan: 224
Vieyra, Rebecca E.: 224
Vignal, Michael: 114, 425
Wagner, DJ: 425, 425
Wan, Tong: 296
Waterson, Alyssa: 425
Watkins, Jacob: 380, 425
Wheeler, Sav: 386, 425
White, Gary: 425
Wieman, Carl: 230
Wilcox, Bethany: 101, 175, 254, 350
Willison, Julia: 391, 425
Wu, Xian: 397, 425, 426
Young, Tamara Gay: 46, 403, 425
Zich, Raymond: 421, 425
Zohrabi Alaee, Dina: 409, 425
Zollman, Dean: 415, 425, 425
Zwickl, Benjamin: 284, 374, 409