ABSTRACT

Title of dissertation: AN ASSESSMENT OF PROFESSIONAL DEVELOPMENT FOR ASTRONOMY AND PHYSICS FACULTY: EXPANDING OUR VISION OF HOW TO SUPPORT FACULTY'S LEARNING ABOUT TEACHING

Alice R. Olmstead, Doctor of Philosophy, 2016

Dissertation directed by: Professor Derek Richardson
Department of Astronomy
and Professor Chandra Turpen
Department of Physics

In this thesis, we will explore approaches to faculty instructional change in astronomy and physics. We primarily focus on professional development (PD) workshops, which are a central mechanism used within our community to help faculty improve their teaching. Although workshops serve a critical role for promoting more equitable instruction, we rarely assess them through careful consideration of how they engage faculty. To encourage a shift towards more reflective, research-informed PD, we developed the Real-Time Professional Development Observation Tool (R-PDOT), to document the form and focus of faculty's engagement during workshops. We then analyze video-recordings of faculty's interactions during the Physics and Astronomy New Faculty Workshop, focusing on instances where faculty might engage in pedagogical sense-making. Finally, we consider insights gained from our own local, team-based effort to improve a course sequence for astronomy majors. We conclude with recommendations for PD leaders and researchers.

AN ASSESSMENT OF PROFESSIONAL DEVELOPMENT FOR ASTRONOMY AND PHYSICS FACULTY: EXPANDING OUR VISION OF HOW TO SUPPORT FACULTY'S LEARNING ABOUT TEACHING

by

Alice Rose Olmstead

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Professor Derek C. Richardson, Co-chair
Professor Chandra A. Turpen, Co-chair
Professor Stuart Vogel
Dr. Melissa Hayes-Gehrke
Professor Edward F. Redish
Professor Edward E. Prather
Professor Andrew Elby, Graduate Dean's Representative

Preface

The majority of this thesis was written in close collaboration with one of my co-advisors, Dr. Chandra Turpen. Some portions of this work have been published as separate works or may be in the future. In particular, Chapter 4 has been submitted to the Physical Review–Physics Education Research and is included here verbatim; Section 5.5.2.1 was closely adapted from our earlier peer-reviewed publication in the 2015 Physics Education Research Conference proceedings [Olmstead and Turpen, 2015]; and Section 5.6 has been incorporated into a submission for the 2016 Physics Education Research Conference proceedings (with Dr. Turpen as first author). Dr. Turpen and I also plan to incorporate 5.4 into a full-length publication to be submitted to the Physical Review–Physics Education Research.

Chapter 6 was written in close collaboration with my other co-advisor, Dr. Derek Richardson, which surrounds a course sequence that he taught as lead instructor. Dr. Richardson's responsibilities in writing this chapter primarily focused on analyzing of student outcomes and drafting descriptions of the course structure. We intend to submit a portion of this chapter for publication in the future.

Acknowledgments

A long thesis deserves an equally long acknowledgments section. The past three years have contained some of the most fulfilling and rewarding experiences of my life, and many people have helped to shape those experiences and support me in doing the work that is represented on these pages.

First off, on the research side, thank you to the New Faculty Workshop team for supporting us in pursuing this research, particularly Bob Hilborn and Stephanie Chasteen. Bob—thank you for letting us be a part of this project, and for giving me access to sit in on workshop sessions from day one. Stephanie—thank you for your enthusiasm and interest surrounding our research, and for lending your expertise to improving the workshop in an immediate sense. Thank you to Colin Wallace for helping me figure out how to get started in my research at the very beginning and believing that I could do this. Thank you to Cassandra Paul, Andrew Reid, and the University of California–Davis Tools for Evidence-based Action team for helping us make our observation tool become real. Thank you also to all of the workshop leaders at the Physics and Astronomy New Faculty Workshop and the Center for Astronomy Education workshop who welcomed me into your sessions. I recognize that leading workshop sessions comes with many challenges, and I have struggled in similar situations myself. Watching you all has given me tremendous insight into how I might strive to act in the future, and without you, none of this research would have been possible. I also want to thank all of the faculty who allowed themselves to be videotaped during workshops and took time out of their overloaded schedules to tell me about their experiences. You all inspired and grounded this work in fundamental ways.

I want to thank all of my science education friends and colleagues who were at the University of Maryland during my time here: Kevin Calabro, Brian Danielak, Ben Dreyfus, Mark Eichenlaub, Andy Elby, Ben Geller, Ayush Gupta, Deborah Hemingway, Hannah Jardine, Brandon Johnson, Vijay Kaul, Eric Kuo, Kim Moore, Gina Quan, Joe Redish, Jen Richards, Vashti Sawtelle, Stephen Secules, Katey Shirey, Erin Sohr, and all of my peers and teachers from my science education classes. Thank you also to my teacher education reading group—Jen, Amy Roberston, and Julia Svoboda-Gouvea. Thank you all for listening to my ideas and telling me about your own research, both of which inspired and challenged me to think in new ways. Thank you in particular to Hannah Jardine for all of your thoughtful contributions to the analysis in Chapter 5, and to Joe and Andy for being on my thesis committee. Joe—thank you for choosing to stay on my committee despite personal hardships that could easily get you off the hook. It means a lot that you want to support me and think this research is so worthwhile.

While I haven't moved to Michigan just yet, I want to thank my postdoc advisors—Andrea Beach and Charles Henderson—who I have begun to work with over these past few months. Thank you for welcoming me into your research family and giving me something to look forward to in the coming two years, and for helping me start to gain a better footing for understanding the landscape of instructional change strategies beyond workshops. Charles—thank you also for agreeing to give a

talk at Maryland over three years ago, thus inadvertently giving the first education research talk I ever attended and helping to set me on this path in the first place.

I'm grateful to have met so many wonderful people in the physics and astronomy education research communities both at Maryland and elsewhere. I want to thank all of the people I've met who are passionate not just about their research but about caring for other people as well, and who are taking strides to do so. You inspire me to try to do good for the world. Thank you also to everyone in the community who has talked and written about work-life balance, which is still a huge struggle for me, and to the people who understand why I find that so challenging. While I imagine that this list will grow longer in the future, for now I will say thank you to Warren Christensen, Kim Coble, Melissa Dancy, Scott Franklin, Ayush, Angie Little, Cassandra Paul, Sam McKagan, Rosemary Russ, Mel Sabella, Vashti, Ellie Sayre, and of course Chandra Turpen and Ed Prather.

I also have many people to thank within the University of Maryland's Astronomy department. First off, thank you to everyone who I have had the pleasure of working with on our course transformation project or teaching with in my own Astro 101 class: Fatima Abdurrahman, Allison Bostrom, Joe DeMartini, Sara Frederick, Emily Garhart, Melissa Hayes-Gehrke, Sarah Scott, Holly Sheets, and Justin Tervala. The moral and intellectual support you all have given me and each other has convinced me that I never want to teach solo. Thank you to the Astronomy faculty for giving me the chance to break the mold and pursue astronomy education research, particularly to Cole Miller for seriously considering my proposal and

advocating for me despite your own reservations about this kind of research, and to Stuart Vogel and Melissa for serving on my committee. My graduate experience was also greatly improved and enhanced by the efforts of my friends and colleagues to make our department more equitable and inclusive over the past few years. I am grateful to have worked alongside so many dedicated and compassionate people. Thank you to everyone who helped me to start and sustain the Astronomy Gentleladies' Network, especially to Sara for taking over my responsibilities in this past semester, and to everyone who has been involved in the leadership of GRAD-MAP, to all the other organizers of the BANG! seminar, to the past and current AstroTerps presidents, and to the Equity and Inclusion Committee members. Stuart—thank you for your incredible leadership of this department in supporting and promoting these efforts. I feel extremely lucky to have been here while you were department chair. Beyond the University of Maryland, thank you to Christina Richey for your leadership in making the Astronomy community more safe, and for taking on so many people's emotional burdens, including some of mine.

Beyond just the issues of academia, thank you to all the friends and family who have stuck with me as I have tried to sort out what I want to do with my life, who have sat with me through emotional breakdowns and celebrated with me when I met with successes. In particular, I want to thank Laura Blecha, Megan DeCesar, Jithin George, Katie Jameson, Kory Kreimeyer, Alison McReynolds, Kim Nguyen, Ed Olmstead, Juliana Olmstead, Jen Richards, Erin Sohr, and Kathryn Williamson. I am lucky to have had all of you in my life.

Last but certainly not least, I want to thank my advisors: Ed Prather, Derek Richardson, and Chandra Turpen. Ed, thank you for being as passionate about professional development workshops as I am. Thank you for having heated arguments with me about workshop design and instructional choices and still giving me a hug afterwards and reassuring me that we're on the same team. Thank you also for asking for my advice and making me feel like my ideas matter, and for speaking so highly of me to other people. Derek, thank you for becoming one of my best friends and letting me hang out at your house almost every week for a semester. Thank you for saying "of course" to every boundary-pushing idea I have had, even when it predictably resulted in you taking on extra work. Thank you for letting me explain so many education research ideas to you over the last few years as I made sense of my own thoughts, thank you for being so willing to let me help shape your instruction from behind the scenes, and thank you for treating me like a colleague without giving any thought to my junior status. Chandra, thank you for absolutely everything. You are an amazing advisor and I could not have done any of this work without you. Thank you for being patient with me when research was tough and I was impatient, and for being so genuinely excited when we articulated meaningful and interesting research claims. Thank you for caring about me as a person and high-fiving me when I started going to therapy, which we both know I needed. Thank you for seeing that all of the work I've done outside of my formal research matters too, and for being a wonderful role model in trying to make our spaces more equitable and inclusive. Thank you for being frustrated when the world is frustrating, and at the same time finding opportunities to celebrate the parts of the world that are good.

Thank you all for reading.

This work is supported by funding from NSF-DUE 1431681, and by the University of Maryland Teaching and Learning Transformation Center Elevate Fellows program.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AER	Astronomy Education Research
Astro 101	Introductory astronomy for non-majors
ASTR120/121	Introductory astronomy for majors at the University of Maryland
CAE	Center for Astronomy Education
COPUS	Classroom Observation Protocol for Undergraduate STEM
E-CLASS	Colorado Learning Attitudes about Science Survey for Experimental Physics
FP	Faculty participant
FPQ	Faculty participant question
Ind	Independent (as in, faculty independent work)
IRR	Inter-rater reliability
IS	Instructional strategy (or strategies)
LA	Learning Assistant
LG	Large group (as in, large group discussion)
LSCI	Light and Spectroscopy Concept Inventory
NFW	Physics and Astronomy New Faculty Workshop
PAER	Physics and Astronomy Education Research
PD	Professional Development
PI	Peer Instruction
PER	Physics Education Research
RBIS	Research-based instructional strategy (or strategies)
R-PDOT	Real-time Professional Development Observation Tool
SG	Small group (as in, small group discussion)
STEM	Science, Technology, Engineering, and Mathematics
TA	Teaching Assistant
TPS	Think-Pair-Share
TLTC	Teaching and Learning Transformation Center
WL	Workshop leader
WLQ	Workshop leader question

Chapter 1: Broad motivation

The pursuit of astronomy and physics knowledge is currently not equally accessible to all. At the professional level, disparities in the participation of people of different gender, race, ethnicity, sexual orientation, ability status, and/or socioeconomic status from the dominant group compared to the U.S. population point to challenges in our field, and in science, technology, engineering, and mathematics (STEM) more broadly [National Science and Technology Council Committee on STEM Education, 2013, Hill et al., 2009, Merner, 2015, Mulvey and Nicholson, 2015]. At the undergraduate level, it is well-documented that women and other underrepresented populations of students disproportionately switch out of STEM majors, even when similarly qualified to (or more qualified than) their peers who do not switch [Hill et al., 2009, Seymour, 2000]. Many signs indicate that the teaching of undergraduate STEM courses has been a major factor that contributes to students' decisions to switch and degrades the experiences of students who do not switch [Seymour, 2000]. A comprehensive, interview-based study in the 1990s revealed that undergraduate students almost universally described "poor teaching" as a factor that negatively impacted their pursuit of STEM careers whether or not they switched to non-STEM fields, and a majority of their other concerns were also tied to classroom instruction to some extent [Seymour, 2000]. Similar, follow-up research reveals that these challenges continue in the present day [Graham et al., 2013]. While undergraduate students' perceptions of "poor teaching" could accurately indicate that instructors simply do not care about teaching, we consider it more likely that their perceptions point to a complex set of factors that prevent instructors from using adequate pedagogical strategies to engage students more equally and to a greater extent. For instance, when instructors heavily rely on traditional instructional strategies such as lecture and calling on individual students to ask or answer questions, a select few students' ideas are often privileged in the classroom, while other students remain silent [Johnson, 2007]. Those students who participate the most vocally are often members of the dominant group, and their interactions with the instructor can subtly discourage other students from participating and persisting in STEM [Johnson, 2007].

Education researchers have developed a set of instructional strategies and principles that can improve student outcomes and experiences beyond what is possible with traditional instructional methods. Astronomy and physics education researchers have developed many curricular materials and classroom practices, broadly referred to as research-based or research-validated instructional strategies (RBIS), that have been shown to positively impact students' conceptual learning and exam performance and decrease failure rates (improve retention) in undergraduate STEM courses [Freeman et al., 2014]. In physics and astronomy, the most commonly used RBIS is *Peer Instruction*, which is a structured way for all students to respond to and discuss challenging conceptual questions during class [Dancy and Henderson, 2010, Mazur, 1997]. Many other strategies promote shifts away from traditional lecture and towards increased student collaboration, and often encompass specific curricular materials that have been developed and tested based on evidence of students' thinking [Prather et al., 2004, Hudgins et al., 2006, McDermott and Shaffer, 1998, Novak et al., 1999, Perkins et al., 2006, Christian and Belloni, 2003, Heller and Hollabaugh, 1992, Heller et al., 1997, Sokoloff and Thornton, 1997, Laws, 1991, Brewe et al., 2010, Redish, 2003]. The education research community at large has also articulated theoretically and empirically-driven principles that can guide instruction. For example, a National Research Council report roughly states: (1) classroom environments should be learner-centered, i.e., focused on building from students' current knowledge, skills, and attitudes; (2) instructors should focus on depth of understanding and go beyond teaching memorizable facts; (3) formative assessments, which provide feedback to both instructors and students, are critical; and (4) instructors should foster a sense of community by working to shape classroom norms and connect to students' experiences beyond the classroom [Bransford et al., 2000]. While RBIS can begin to enable instructors to follow these ideals by giving them useful instructional tools, these research-based principles can guide the design of equitable and engaging classroom environments more generally.

Empirical studies have demonstrated that the use of RBIS in the classroom can contribute to improved outcomes for all students relative to traditional instruction. In particular, some researchers have drawn attention to equity issues by looking for correlations between quantifiable student outcomes (e.g., those measured by research-based assessment instruments) and demographic variables, and find promising signs. For example, a large, national study of students' conceptual learning in introductory astronomy for non-majors showed that students with a variety of demographic characteristics benefited equally from the incorporation of a particular set of RBIS into the classroom: gains in conceptual understanding did not correlate with students' gender, ethnicity, academic preparation, incoming GPA, or primary language [Rudolph et al., 2010]. Instead, the amount of interactive engagement was the only measured variable that was consequential to students' conceptual learning. [Prather and Brissenden, 2008] use these results to illustrate the potential RBIS to help all students (as opposed to helping some students at the expense of others). Research on the implementation of another RBIS at a primarily Hispanic-serving institution also presents a highly promising case. Physics education researchers at Florida International University have demonstrated that implementing a RBIS called *Modeling Instruction* in introductory physics courses has led to a variety of positive student outcomes, including improvements in students' conceptual understanding, success in the course overall, attitudes towards physics, and self-efficacy compared to lecture-based courses at the same institution [Brewe et al., 2010, Traxler and Brewe, 2015, Sawtelle et al., 2012].

Otherwise successful RBIS implementation might not produce equitable outcomes by all metrics, however, and it is unclear which metrics we should rely on if we are interested in equity issues. In particular, researchers have suggested that it may be insufficient to only look at students' conceptual learning in diagnosing how students from underrepresented groups benefit from research-based instruction [Brewe et al., 2010, Sawtelle et al., 2012, Hazari et al., 2007]. Revisiting the case of Modeling Instruction from above, when Brewe et al. [2010] looked at shifts in students' conceptual understanding between the start and end of the course, while all students learned more than students in the comparison lecture-based course and the ethnicity gap in conceptual understanding remained the same, the gender gap actually widened, which was unexpected. At the same time, the odds of passing the course, used as a proxy for students' success overall, revealed no gender or ethnicity gap, and also showed a dramatic improvement over traditional lecture for all students. Sawtelle et al. [2012] argue that this provides evidence that conceptual inventories are not measuring important aspects of students' experiences. Instead, the authors show how students' self-efficacy was a good predictor of passing this physics course, and that the frequent opportunities for students to share in their peers' successes afforded by this instructional approach mattered most for improving female students' self-efficacy [Sawtelle et al., 2012]. This demonstrates how looking beyond conceptual learning gains may be important for understanding equity in course outcomes.

These examples also bring up some murkiness in how "equity" should be defined. While it is clear that an increased gender gap in metrics of interest would not, by itself, indicate an equitable outcome (since the end result appears to be increased disparity), it is unclear whether it should be considered more desirable for all students to gain the same amount, or for initial gaps to be reduced, which requires that students who are less prepared initially benefit more than students who are already privileged coming into the course [Rodriguez et al., 2012]. Rodriguez et al. [2012] describe these two kinds of outcomes as "equity of fairness" and "equity of parity", respectively: the first case is "fair" because no student group gains more than another, while the second case represents parity because all students would leave equally prepared, i.e., all students finish the course with equivalent disciplinary knowledge or beliefs. Outcomes consistent with "equity of fairness" but not "equity of parity" do not seem to fully support an ultimate goal of equal participation in physics and astronomy for all, since some populations of students seem to leave the class less prepared to continue on in our discipline than others [Kost et al., 2009]. But we also recognize that "equity of parity" is an ambitious goal that may not be achievable in a single semester and without increased supports for students outside of the classroom, and that even students who score less well on particular assessments could leave equally or more prepared relative to their classmates in other ways.

There is a wide range in the outcomes of using various research-based instructional strategies, and some results have been less positive than others. For one, not all interactive engagement leads to increased conceptual learning. Variations in the implementation of RBIS can be highly consequential to how much students learn (in aggregate), and some use of interactive engagement produces similar (though not worse) learning outcomes relative to traditional lecture [Hake, 1998, Prather and Brissenden, 2008, Turpen and Finkelstein, 2010]. Important aspects of RBIS are often underspecified by the developers [Turpen and Finkelstein, 2009, 2010], and could be impossible to fully prescribe when considering the importance of genuinely encouraging instructors to be responsive to students' ideas and behaviors [Robertson, 2015]. Furthermore, while all RBIS draw on some of the research-based principles described above, they vary in the extent to which they are aligned with the full breadth of these educational ideals [Dancy and Henderson, 2007]. Because of this, the exact choice of RBIS will likely promote or constrain the potential for favorable student outcomes, which may or may not be captured with current assessment instruments [Dancy and Henderson, 2007]. When it comes to equity concerns, shifts in students' attitudes towards learning physics and astronomy (or other non-cognitive outcomes like self-efficacy, as above) may better indicate the nature of their experiences and predict whether or not they will choose to persist in the field than conceptual learning gains alone. Unfortunately, the positive attitude shift Traxler and Brewe [2015] observed in Modeling Instruction may be a rare exemplar: in physics courses that use different instructional methods (research-based or not), student attitudes or beliefs about physics either stay the same or worsen when they take physics courses, and it seems that physics instruction tends to preferentially encourage students who already have expert-like beliefs before they enter college to continue on as physics majors [Perkin and Gratny, 2010, Madsen et al., 2015]. However, systematic variations in these outcomes by student population have not been explored, and there is much that the community does not know about how instruction impacts student attitudes and the implications for equity considerations (e.g., which students persist in physics and astronomy) [Madsen et al., 2015].

We also know that marginalization of underrepresented students can continue even when RBIS are incorporated into classroom instruction. This is in part because the use of RBIS does not automatically reduce implicit biases that influence how instructors interact with students, and how students interact with their peers during collaborative learning [Carlone, 2004, Secules et al., 2016, Barron, 2003, Esmonde, 2009, Steele, 2010]. Similarly, not all students experience their race, gender, ethnicity, etc. in the same way, and therefore will not experience the classroom environment in the same way [Traxler et al., 2015, Parks and Schmeichel, 2012]—students with shared demographic characteristics may experience instructional innovations differently. Moreover, many existing RBIS (particularly in physics) were developed at traditional research institutions with primarily White students, and may require modification to best suit the strengths and needs of underrepresented student populations, particularly students at minority-serving institutions [Sabella, 2002, Sabella et al., 2008]. Thus, some instructional tools that could better help students in underrepresented groups may not yet be well-developed, and more work at a variety of institution types is likely needed.

The potential for these kinds of strategies to improve undergraduate astronomy and physics education, as well as the need to collectively learn more about how different instructional strategies work in a variety of contexts, presents a strong case for involving faculty as partners in this process—both in teaching them what we know now and in working with them as we all continue to explore ways to design better and more equitable classroom environments. Faculty have significant freedom in how they teach [Kezar, 2001], and can directly and immediately influence their students' experiences. Moreover, faculty can play a significant role in influencing the instruction of their peers and advocating for change within their institutions [Dancy et al., 2010, Kezar, 2001, Chasteen et al., 2011]. Because of this, faculty have substantial (though not unlimited) power to improve undergraduate instruction, and working with them presents a critical lever for increasing equity in STEM.

At the same time, we cannot expect faculty to easily and unproblematically pick up instructional methods that are different from traditional instruction. The term "traditional" holds weight here: faculty have learned much of what they know about teaching through observing their own teachers, and thus have a natural inclination to teach how they were taught [Lortie and Clement, 1975]. Because of this, faculty need support in making the kinds of changes that education researchers have begun to envision, and we should not expect the change process to be entirely smooth.

In the current landscape, research-based strategies and principles are not broadly used in undergraduate physics and astronomy classrooms across the country, and faculty's interactions with education researchers are often limited, as is true in other STEM fields [Henderson et al., 2012, Borrego et al., 2010, Macdonald et al., 2005, Henderson and Dancy, 2008, Dancy et al., 2010, President's Council of Advisors on Science and Technology, 2012]. For instance, a national survey showed that although a vast majority of physics and astronomy faculty are aware of RBIS (88%) and most become motivated to try and do try RBIS in their classrooms (72%), about a third of these faculty discontinue their use after a single semester [Henderson et al., 2012]. More detailed research reveals that these faculty perceive many barriers to using RBIS in their local contexts [Turpen et al., 2016], which can contribute to their decisions to discontinue use of these strategies. Similarly, some faculty who claim to be implementing RBIS sometimes modify them in ways that bring them farther from the education research ideals the designers were striving to foster, and thus are less likely to achieve the same positive outcomes as the developers [Dancy et al., 2016].

These challenges motivate us to critically examine how we teach faculty about research-based instructional innovations. Even though we may not be able to give faculty a fully prescribed, foolproof way to reach all their students, if we guide them to find value in existing instructional tools, they might be able to envision how research-based instruction fits within their local contexts and build from RBIS to make productive instructional choices that go beyond what we have specified or named for them. Faculty likely already have some rich knowledge of their student body, their departments, and their institutions, and they are uniquely positioned to continue to learn about these groups and the individuals within them. This local expertise—both what faculty know now and what they could learn through future local interactions—is essential for developing and enacting plans of action and sustaining instructional change. But in order for faculty to be well-prepared to make wise instructional decisions based on education researchers' recommendations and local contextual knowledge, they need scaffolded opportunities to make sense of existing instructional innovations and student reasoning for themselves. This thesis will explore what these opportunities could and do look like, and provide recommendations for how the physics and astronomy education research communities could use our existing resources to promote more sustainable and progressive instructional change.

Chapter 2: Empirical, theoretical, and analytical frameworks

2.1 Empirical results on instructional change

Henderson et al. [2011] created a comprehensive framework to characterize others' efforts to improve undergraduate STEM instruction. The categories defined in this work have since become deeply integrated into the language used by higher education researchers and other proponents of instructional change. Here, we begin to situate the instructional change efforts explored in this thesis within prior empirical results by explaining Henderson et al. [2011]'s framework and considering where the efforts we will study seem to fall within it.

In brief, Henderson et al. [2011] looked across 191 published studies and found that instructional changes strategies can be meaningfully characterized by whether they target individuals or environments/groups, and whether they target prescriptive or emergent outcomes. These two dimensions define four change approaches: *disseminating curriculum and pedagogy* (individual/prescribed); *developing reflective teachers* (individual/emergent); *enacting policy* (group/prescribed); and *developing shared vision* (group/emergent), as depicted in Figure 2.1.

We can begin to understand each of these approaches by considering examples within physics and astronomy education research (PAER)/physics education

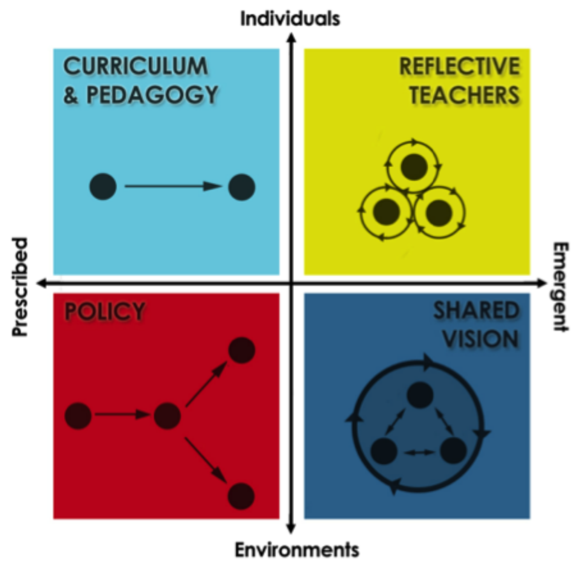

Figure 2.1: A schematic view of the four change strategies for improving undergraduate STEM education, adapted from Figure 1 of Henderson et al. [2011] (color figure taken from https://wmich.edu/changeresearch).

research (PER), as well as examples at the University of Maryland and similar institutions. Starting in the upper left quadrant, *disseminating curriculum and pedagogy* is a highly common approach taken within PAER. Many research-based instructional strategies for physics and astronomy have been developed, tested, and refined by a relatively small group of education researchers and/or instructors, often at one or a few institutions, with the intention of encouraging other instructors to implement these strategies in highly similar ways. Popular RBIS like this within astronomy and physics include Peer Instruction (PI) [Mazur, 1997] or Think-Pair-Share (TPS) [Prather et al., 2009], Tutorials in Introductory Physics [McDermott and Shaffer, 1998], Lecture-Tutorials in Introductory Astronomy [Prather et al., 2004], Collaborative Ranking Tasks for introductory Astronomy [Hudgins et al., 2006], Just-in-Time Teaching [Novak et al., 1999], and several others listed in Dancy and Henderson [2010]. Professional development workshops are frequently used as a mechanism for disseminating these strategies: the Physics and Astronomy New Faculty Workshop [Henderson, 2008], the Center for Astronomy Education Tier I Teaching Excellence Workshop [Prather et al., 2009], and one-day workshops at American Association of Physics Teachers meetings are all places where this kind of change strategy is enacted. Similarly, static online resources like websites (e.g., PhysPort.org) or YouTube videos (e.g., Eric Mazur's "Confessions of a Converted Lecturer") are mechanisms for encouraging individual physics and astronomy faculty to implement existing, pre-defined instructional strategies.

Change efforts that fit within the *developing reflective teachers* quadrant aim to cultivate pedagogical expertise in individual instructors by guiding them to explore a variety of instructional strategies and reflect on their own instruction. The leaders of these change efforts focus on helping instructors to identify and learn from meaningful classroom events and assessments. Instructors are supported by facilitators or their peers as they try out existing instructional strategies or create new ones, but their choice of instructional strategies is not constrained by the facilitators. Centers for teaching excellence, including the Teaching and Learning Transformation Center at the University of Maryland, often take this approach by assembling and facilitating faculty learning communities on their campuses. In the typical model for faculty learning communities, groups of 8-12 instructors from a variety of disciplines meet regularly to discuss instruction over a period of about a year [Beach and Cox, 2009]. Many of these faculty learning communities encourage faculty to think about teaching and learning in a scholarly way by carrying out small-scale research investigations in their own classrooms [Hutchings and Shulman, 1999]. Within the physics and astronomy communities, the New Faculty Workshop recently began encouraging participants to join Faculty Online Learning Communities following the workshop, which serve a similar function to traditional faculty learning communities by supporting faculty in refining their instruction and pedagogical thinking, but allow participants to discuss shared disciplinary ideas and materials and be geographically dispersed [Rundquist et al., 2015]. The Periscope project, which is an online collection of video excerpts from undergraduate physics classes paired with open-ended discussion prompts, is another example of an effort to develop reflective teachers within PER [Scherr and Goertzen, 2015].

Change efforts focused on *enacting policy* tend to be prescriptive at the institutional level. These efforts almost always involve university administrators (provosts, deans, etc.) in the project leadership, but can involve department chairs and other faculty as well. One common approach within this category is to work on changing the requirements for promotion and tenure to support faculty who are experimenting with research-based teaching. Institution-level changes can help faculty to feel less constrained in what they can safely try in the classroom: faculty risk lower student evaluations as a result of trying out instructional strategies that demand more from students and are unfamiliar to both instructors and students [Foote et al., 2016, Seidel and Tanner, 2013], yet student evaluations are often leaned on heavily in assessing teaching for tenure and promotion cases [Henderson et al., 2014]. This change approach seems particularly fruitful when many stakeholders are included in policy discussions. For example, at the University of Colorado-Boulder, a team of physics education researchers, institutional administrators, and a select number of faculty representatives are working to develop a "teaching quality framework" that will be used to assess teaching at their institution and will align with other simultaneous change efforts [Corbo et al., 2016].

Change efforts characterized by *developing shared vision* also involve teams or groups of stakeholders, like *enacting policy*, and target emergent outcomes, like *developing reflective teachers*. Unlike interdisciplinary faculty learning communities, instructors who participate in these efforts must work towards developing a shared product, i.e., they must come to consensus about instructional decisions in order for the change initiative to move forward, and there is often a greater variety in who participates in these efforts. In particular instructional development teams often form within departments or are intended to comprise entire departments, and can include undergraduate students, graduate students, postdoctoral researchers, tenure and non-tenure track faculty, and science education specialists in making joint decisions. The Science Education Initiative at the University of Colorado-Boulder and the University of British Columbia has supported many change efforts that can be characterized in this way [Corbo et al., 2016, Chasteen et al., 2011, 2015]; for example, the Boulder Physics department worked with Science Education Initiative staff over multiple semesters to transform all of their upper-level physics courses for undergraduate majors. While the presence of a local PER group may have shifted ownership away from non-PER faculty to an extent, even in this case, the process enabled non-PER faculty to develop ownership over the proposed changes in multiple ways: faculty working groups developed learning goals for the target courses, change agents investigated student learning difficulties and shared these with faculty to inform their instructional choices, and any instructor who taught these transformed courses had agency to modify and add to the initial course materials in subsequent semesters, and many did so [Chasteen et al., 2015].

2.1.1 Locating thesis components within this empirical framework

This thesis will investigate two instructional change efforts that span multiple quadrants of Henderson et al. [2011]: a national, teaching-focused professional development workshop for physics faculty and a local, team-based effort to transform an introductory astronomy course sequence for majors. We now describe how these two efforts fit within this framework in the text that follows.

The setting for the majority of this thesis research is the Physics and Astronomy New Faculty Workshop (NFW) [Henderson, 2008], which primarily fits into the "disseminating curriculum and pedagogy" approach. The NFW has been a critical mechanism for disseminating RBIS in the physics and astronomy community: attending the NFW is the strongest predictor of whether faculty are aware of and have tried RBIS in their classrooms [Henderson et al., 2012]. In locating it in this top left quadrant, we note that a majority of the NFW sessions have been focused on disseminating specific RBIS, and that often only one faculty member from each institution attends the NFW at a particular time, indicating a focus on individual outcomes [Henderson, 2008]. These sessions are often led by original RBIS developers, which could lead to a high degree of prescriptiveness since the developers have not needed to adapt these strategies and may be less aware of what modifications may be appropriate than a secondary implementer.

At the same time, because there are some known drawbacks to a pure dissemination approach, we are interested in whether and how the NFW might deviate from the characteristics outlined for this category. Specifically, we think that it is important to diagnose whether and to what extent the NFW promotes instructional change in ways that go beyond encouraging individual faculty to adopt existing strategies. A purely prescriptive approach (simply telling faculty about the details of existing innovations) implicitly assumes that faculty will adopt RBIS as-is, but in reality faculty want or need to adapt and modify these strategies to fit their local contexts, i.e., disseminators' goals often do not match faculty's needs and expectations [Henderson, 2008, Turpen et al., 2016, Dancy et al., 2016]. This disconnect likely contributes to faculty's struggles to implement RBIS effectively in their classrooms [Turpen et al., 2016]. Related to this, we know that changing instruction is a long-term process, and as [Henderson et al., 2012] and many designers of K-12 teacher professional development have suggested, instructors need long-term support that workshops alone cannot provide [Loucks-Horsley et al., 2009, Desimone et al., 2002, Garet et al., 2001, Darling-Hammond et al., 2009, Wilson, 2013]. Ideally, workshops could catalyze faculty's continued engagement with the PAER community, but portraying existing instructional strategies as rigid or inflexible (at the NFW or elsewhere) can lead to negative repercussions in faculty's attitudes towards education research that discourage these continuing interactions. For example, some physics faculty who are otherwise dedicated to education perceive physics education researchers to be dogmatic, and seem to have pulled away from using RBIS and engaging with the education research community in part as a result [Henderson and Dancy, 2008].

Since there have been no prior in situ investigations of the NFW and the structure of the workshop has been shifting over time, it is plausible to us that the current NFW could target some emergent outcomes and thus ameliorate these challenges to an extent. Within sessions, workshop leaders might support faculty in considering how to modify and adapt of RBIS for their local contexts instead of prescribing a single approach. This would be an intrinsically emergent orientation to introducing RBIS even if the primary goal could still be described as dissemination. It seems plausible to us that faculty will be more likely to persist at implementing active learning strategies in the long-term if their early professional development experiences are responsive to their needs and prior experiences. If workshop leaders design and facilitate opportunities for faculty to engage in open discussions about instruction during the NFW, these discussions could be similar to discussions we might expect to observe within developing reflective teachers or shared vision efforts. Eliciting and listening to faculty's contributions during these discussions could create opportunities for workshop leaders to identify alignment with faculty's ideas and experiences. This kind of workshop facilitation could build up positive faculty perceptions of PAER, and help faculty to refine or stabilize their ways of thinking about teaching and learning. Lastly, although the NFW leaves few opportunities for long-term collaborative work, NFW sessions might explicitly encourage faculty to seek out opportunities for continued engagement with other educators at their home institutions and help them generate concrete plans for how they might do so.

Indeed, recent changes to the NFW design incorporate an optional component—faculty online learning communities, mentioned above—which do fall neatly into the "developing reflective teachers" approach, so we know in theory that the NFW is followed by an emergent change approach for some participants.

Looking to the second component of this dissertation, the local course transformation work is well-described by a "developing shared vision" change approach. In this course transformation effort, multiple instructors have contributed to the development of a shared product—the instruction of the course itself—both through behind-the-scenes design of the course structure, learning goals, and tasks for students, and through the teaching of the course itself. While the outcomes of this effort can easily be thought of as emergent, the instructional changes were strongly supported by previous dissemination of curriculum and pedagogy within PAER. Two of the team members (Alice and Melissa) have participated in multiple professional development workshops and gained extensive awareness of research-based instructional strategies and materials prior to this effort. Because of this, even though none of the change agents involved in the project had dissemination as a singular goal, many existing RBIS were incorporated into the course sequence with various degrees of adaptation, modification, and reinvention [Henderson and Dancy, 2008]. Several of the team leaders also participated in a local faculty learning community that closely tracked the "developing reflective teachers" model, which pulls this effort away from being exclusively shared vision. We suspect that these deviations strengthened the current form of instruction in this course sequence.

2.2 Theoretical approaches to instructional change

2.2.1 What is a theoretical framework?

Because the notion of choosing a theoretical approach may feel unfamiliar to some of our readers, I will take a moment to explain what this means before proceeding. Colloquially, a theoretical approach or *theoretical framework* can be thought of as a way of looking at the world. It can greatly benefit readers when researchers articulate their theoretical stance before explaining their analysis and results, because the reader can then try to orient to the world in similar ways and thus make appropriate judgments about what has been done and what has been suggested about what to do next. In some ways, the idea of choosing a theoretical approach is unique to social science research: the things that we want to model (people's ideas and interactions) are more complex and contingent than what one would model in traditional science research (physical objects and processes) and therefore there is greater flexibility in what theoretical approaches we can choose to draw from without being unequivocally "wrong." At the same time, I would argue that astronomers are not entirely unfamiliar with this kind of process in their own research. In computational astrophysics, for example, different numerical methods have different affordances and drawbacks—one method may be more accurate but less efficient (and thus more expensive) than another—meaning that there often is not a single correct way to model a particular physical situation. Instead, researchers must make judgments about what they want to accomplish and whether the potential drawbacks are tenable. Thus one might begin to conceptualize choosing a theoretical framework as similar to choosing a methodological approach in astronomy research.

Despite making this comparison, I do not mean to suggest that these two processes are equivalent. Instead, I would hope that this comparison might both help to ground more scientifically-oriented readers' understanding of education research theory and reveal why the choice of theoretical approach in education research can be significantly more complex. In the example above, the choice astronomy researchers need to make is fairly constrained: there are only a few options, and the decision has a fairly limited influence on the design of the study and the kinds of results that might emerge. In education research, there are many theoretical frameworks that one could draw on, each with a wide variety of affordances and limitations that are inherently connected to various perspectives about the nature and purpose of learning. There is often space to reinterpret these theories in various ways, and researchers often combine multiple theoretical perspectives to create a more complete and coherent picture that matches their stance towards education.

In this thesis, articulating a theoretical approach will help us to establish what we think is important about learning and to situate our assumptions within an existing body of scholarly work. Our theoretical perspective also influences the mechanics of our study by coloring our decisions about what kind of data to centrally analyze, and what we might count as evidence of learning within this analysis. For example, a theoretical approach that defines learning as an individual process which occurs solely within the learner's head would be well-matched to centrally analyzing one-on-one interview data, while a theoretical approach that defines learning as a more distributed and situated process that involves interactions with other people and materials would be well-matched to centrally analyzing video data of collaborative work. (Indeed, our thinking around the utility of these two kinds of theoretical frameworks does importantly color how we justify our analysis, as we will discuss below.) Because our theoretical perspective influences our study design and analysis, it naturally influences our conclusions and recommendations as well. For all of these reasons, we will take time to introduce and establish our stance towards some existing theoretical frameworks in the section that follows.

2.2.2 Limitations of prior models of faculty thinking

Given that we endeavor to study faculty's learning about teaching, here, we will establish which theoretical frameworks most closely match our orientation to faculty's thinking. However, in exploring past research, we find much of the prior theoretical modeling of faculty's thinking is misaligned with our purposes. Typical literature on this topic has focused on common shortcomings in faculty's knowledge about teaching and learning, and has not considered the potential effects of the situation that faculty are embedded in [Kember, 1997, Samuelowicz and Bain, 2001, Dall'Alba, 1991, Trigwell and Prosser, 1996]. Many research studies have sought to characterize faculty's conceptions about teaching, under the fundamental assumption that every instructor has a single, coherent set of conceptions about teaching and learning that constrains their instruction [Kember, 1997]. This theoretical framework has led researchers to ask faculty broad, decontextualized questions about the nature of teaching and learning (e.g., "What is teaching?" [Trigwell and Prosser, 1996]), and resulted in similar conclusions across many initial, independent studies. In particular, researchers identify and label university instructors' conceptions in highly similar ways, typically as a linear progression of conceptual stages ranging from highly teacher-centered to highly student-centered. These researchers almost universally agree that more student-centered conceptions are superior to more teacher-centered conceptions, and that many instructors hold these teacher-centered conceptions. Statements that imply hierarchy were often quite explicit: for example, Trigwell and Prosser [1996] denote instructors' different conceptions as "levels", and describe more student-centered conceptions as "more sophisticated" and "higher up [on] the hierarchy", while teacher-centered conceptions are described as "less sophisticated", "limiting", and "low-level approaches to teaching."

These findings demonstrate that a conceptions-oriented theoretical framework naturally leads researchers to articulate a deficit model of many faculty's thinking about teaching. Because this theoretical approach is so common in the literature, we suspect that it informs and/or reflects underlying assumptions behind some of the practical approaches to instructional change illustrated in Section 2.1. In particular, the most prescriptive, individually-focused instructional change efforts seem to implicitly align with a deficit model: change agents who exclusively focus on prescribing the desired final form of faculty's instruction without conceding the possibility of productive adaptations or modifications likely assume that faculty are not capable of adequately responding to emergent instructional issues. While conceptions-oriented researchers tend to recommend other PD approaches, such as focusing on improving faculty's conceptions of teaching as a precursor to introducing them to new instructional strategies [Kember, 1997, Trigwell and Prosser, 1996], we can see how change agents who both want to promote the use of specific research-based materials or pedagogy and hold a deficit view of faculty's thinking would default to providing faculty with a detailed roadmap for how to proceed. Alternatively, change agents who closely follow the recommendations of this literature would likely consider it a necessary step to elicit, confront, and replace or resolve any "low-level" faculty conceptions through PD.

Conceptions-oriented studies of faculty's thinking mirror studies of students' conceptions about disciplinary ideas [Kember, 1997], and criticisms of a conceptions-oriented approach to understanding student thinking are apt here as well. More specifically, both the notion that an individual faculty member has a single, coherent mental model of teaching and learning and the focus on drawing attention to certain "limiting" or problematic models, are closely analogous to a focus on identifying student "misconceptions." Smith III et al. [1994] argue that while misconceptions research can be leveraged in productive ways, the complementary stance that students' ideas should be erased and replaced goes against a central principle of education research, constructivism, which posits that learners develop more advanced knowledge by building on prior knowledge [Bransford et al., 2000, Redish, 1994]. Moreover, the authors argue that a focus on misconceptions often leads to an instructional approach that "essentially denies the validity of students' ideas" and "communicates to students that their specific conceptions and their general efforts to understand are fundamentally flawed" [Smith III et al., 1994, p. 126]. It is an oversimplification, however, to assume that learners—either students or instructors—have robust, unitary conceptions that they apply in all situations [Hammer et al., 2005, Disessa and Sherin, 1998], and thus the complexity and diversity of learners' ideas is typically not well-captured by these prior studies. Instead, we choose to draw from other theoretical frameworks that can account for the potential nuances of faculty's thinking in our analysis and help change agents to think about how to build from instructors' prior knowledge.

2.2.3 Re-orienting to faculty thinking

We now describe our starting assumptions about how faculty learn, our interpretation of how some common theoretical frameworks align with these assumptions and support our overall stance towards promoting instructional change, and our use of these theoretical ideas within this thesis.

Foundationally, we draw on theoretical frameworks that naturally allow us to see the value in faculty's ideas. We agree with scholars who have interpreted basic constructivist principles to argue that learners' initial ideas, even when not appearing to be canonically "correct", are critical to learning and therefore intrinsically valuable [Duckworth, 1996, Hammer, 2000]. More complex ideas necessarily build from existing ideas, or as Duckworth [1996] states:

> Intelligence cannot develop without matter to think about. Making new connections depends on knowing enough about something in the first

> place to provide a basis for thinking of other things to do—of other questions to ask—that demand more complex connections in order to make sense. The more ideas about something people already have at their disposal, the more new ideas occur and the more they can coordinate to build up still more complicated schemes. [Duckworth, 1996, p. 14]

Given this stance, it is perhaps unsurprising that Duckworth [1996] characterizes intellectual development by "the having of wonderful ideas": because learners' initial ideas are central to the building up of more complex and robust knowledge, any potentially relevant ideas can (and should) be seen in a positive light. Saying learners have "wonderful ideas" is not to say that learners will automatically draw on their most productive ideas for a particular situation. Instead, researchers like Hammer et al. [2005] argue that while learners can hold ideas that others would identify as "misconceptions," these misconceptions are based on "small" ideas that are productive in other contexts, and it is neither realistic nor desirable to try to replace these ideas entirely. This way of thinking has motivated researchers to pay careful attention to the contexts in which learners' ideas are coming up, and when "misconceptions" arise, to search for strategic, alternative situations that are likely to cue up ideas that are more relevant to the desired situation [Hammer et al., 2005]. When we can conceive of these alternative situations that allow learners' to extend their thinking in small steps to establish more complex knowledge, instead of drastically changing their existing mental models without these footholds, learning becomes easier [Redish, 1994].

More formally, Hammer et al. [2005]'s conceptualization of learners' ideas as *resources* is a theoretical cornerstone that we will use in building up our picture of faculty's thinking and learning. A key characteristic of resources is that they will be cued up or activated in ways that match the learner's sense of what is happening at that moment (the perceived situation or context) and the kinds of ideas they typically think with in those situations. The process that determines which resources are activated, i.e., the learner's answer to "what is it that's going on here?", is called *framing* [Hammer et al., 2005, p. 9; MacLachlan and Reid, 1994;Tannen, 1993]. Hammer et al. [2005] argue that these finer-grained ideas, not conceptions, are the cognitive elements that researchers and teachers should focus on. A resources-oriented theoretical framework describes learners' existing knowledge as divided into potentially valuable, potentially uncoordinated pieces that can be reorganized and used appropriately. As learning occurs, the reorganization of learners' ideas might develop stability such that a particular context or framing activates more appropriate, more coordinated, or more complex sets of resources over time. In this way, the central challenge in designing learning environments can be thought of as a question of how to activate and reorganize learners' existing ideas to meet various instructional goals, instead of a question of how to elicit, confront, and replace their misconceptions.

Applying this theoretical stance towards learners' ideas to faculty can help us to think about instructors' knowledge in ways that will not lead us to identify deficiencies when instructors place significant value in "traditional" teaching methods. While it has not typically been done in the literature, we are not the first researchers to suggest that there is value in this theoretical approach for understanding teachers' thinking. Goertzen et al. [2010], Harlow et al. [2013], and Markauskaite and Goodyear [2014] all explore potential instructor resources, and consider how these resources might be elicited and applied in useful ways that support research-based teaching. As we move towards articulating our own interpretations of the utility of this theoretical framework for describing faculty's thinking and designing PD, we first summarize these three studies and consider the authors' recommendations for PD leaders in the following two paragraphs.

Goertzen et al. [2010] and Harlow et al. [2013] identify specific ideas that are salient to instructors and consider how these specific ideas might support instructors in using research-based instructional methods. In both cases, the instructors' ideas or actions often seem like they are misaligned with prescribed, research-based strategies, but the authors argue that natural paths towards alignment do exist and could be supported by the resources they identify. For instance, in Goertzen et al. [2010], Alan, a Teaching Assistant who facilitated students' group work on Tutorials in Introductory Physics (a popular RBIS) [McDermott and Shaffer, 1998], would often quickly tell students the correct answer instead of giving them space to discover the answers for themselves by offering gentle guidance and asking Socratic questions (as the designers intended). Goertzen et al. [2010] find that Alan's intent was to help students, which they identify as a resource that might support this instructional method if they talked with Alan about what helping students could mean (e.g., what it would look like to help students become better learners).

Harlow et al. [2013] identify a similar potential barrier to pre-service teachers (i.e., teachers-in-training) applying the idea that teachers should guide students instead of telling them answers, worrying that students could become lost or frustrated and might leave their classes with incorrect ideas about science. Instead of identifying an underlying value that could be productive (like Alan's desire to help students), Harlow et al. [2013] label the idea "guiding students is less certain than telling them (the right answer)" as a resource. They note that while this idea sometimes created resistance to research-based recommendations, the pre-service teachers also used it to argue for the benefits of guiding over telling, stating that giving students the answers quickly might condition them to think less because they would anticipate the teacher's willingness to provide the appropriate reasoning for them.

Thus, in both of these examples, a given resource could motivate or discourage teachers from following the research recommendation that questioning and guiding students is typically preferable to telling them answers. Because these resources may not naturally arise in ways that support the use of RBIS, and because these resources are highly salient to instructors, it may be useful for PD leaders to intentionally elicit these ideas and explicitly encourage instructors to consider in what context these resources may or may not be productive for instruction.

Markauskaite and Goodyear [2014] also claim that like student knowledge, teacher knowledge is comprised of many small ideas that are tied to particular contexts, and that teachers can hold conflicting ideas without issue. Instead of identifying specific resources, they identify general frames of reference and sources of knowledge that teachers could draw from. In particular, they suggest that teachers make different judgments about how to teach depending on whether they are orienting to their disciplinary content knowledge, their students' needs and interests, or the affordances and constraints of their instructional environment, or other "frames of reference." They argue that teachers' ability to determine which frame of reference will be the most appropriate for a particular situation is a key part of their pedagogical expertise, and one that PD leaders should focus on helping them to develop.

For us, the assumption that faculty hold many potentially productive ideas that may be cued up in different contexts, and that PD leaders could help faculty to refine and reorganize these ideas, centrally informs how we orient to faculty's thinking and learning. We agree that instructors' "resources" will not be inherently productive towards any particular goal, but could be taken up in ways that support research-based instruction. We also agree that instructors might orient to teaching in multiple ways, and that different kinds of ideas could emerge depending on the framing that instructors adopt in that moment. As we begin to consider how a resources-oriented theoretical framework might influence our interpretations of how faculty act during workshops, as well as how workshops could be re-envisioned, we find it useful to speculate about what specific situations (i.e., framings) might naturally cue up potentially fruitful ideas for faculty's instruction. Faculty could have a variety of ways of organizing activities for the learning of others: for example, many faculty will mentor students in research, raise children, give talks for different audiences, and address their colleagues' disciplinary questions and confusion during informal conversations. Each of these contexts might cue up different cognitive resources that could be productively applied towards reasoning about instructional choices. Lending weight to our speculation, Oleson and Hora [2013] demonstrated that faculty use of some of these sources of knowledge when making decisions about how to teach, in addition to thinking about their prior experiences in the classroom as both instructors and students. Thus, looking for the complexity and richness of faculty's existing ideas can provide a useful lens into their current instructional practices, and can help us to respectfully understand (as researchers) and try to build from (as PD leaders) faculty's ideas about teaching and learning.

Resources and framing might have explanatory power for understanding faculty's engagement in PD workshops specifically. For example, consider a common scenario where faculty are asked to collaborate on a research-based science task during a PD workshop, taking on the role of pseudo-students [Council of Scientific Society Presidents, 2012, Prather et al., 2009]. We suggest that faculty could orient to this task in (at least) two distinct ways, and that these distinct orientations would likely influence what they say and do. In one case, providing faculty with a task designed for students might cue up faculty's ideas about how they were historically expected to act as students. What faculty would think, say, and do in workshops could thus reflect what they would think, say, and do in traditional school environments [Ball and Cohen, 1999, Lortie and Clement, 1975], such as trying to follow the classroom "rules", articulating and writing down plausibly correct answers without exploring alternatives, and individual faculty members prioritizing their own understanding over their peers' understanding (driven by competition) [Pope, 2001, Jimenez-Aleixandre et al., 2000, Lemke, 1989]. Alternatively, faculty might orient to the task as an instance of figuring something out with their peers, more closely mirroring how they might solve authentic research problems with their colleagues now. In this case, we might expect faculty to think, talk, and act in different ways, such as trying to make sense of the science content at hand, taking the time to fully articulate their reasoning and explore alternatives before moving on, and prioritizing the development of shared understanding with their peers even if they are satisfied with their own reasoning (driven by collaboration). As we will substantiate with evidence in Chapter 5, both of these modes of faculty interaction occur within PD workshops, and this theoretical framework can help us start to understand why.

While considering faculty's cognitive resources can generate insights into how PD could support faculty's learning that we would be unable to identify when assuming faculty have stable conceptions (and misconceptions) about teaching and learning, we might still overlook key factors likely to influence the form of faculty's future instruction if we restrict our view of faculty's learning to a process that occurs in individual's heads. In particular, because we ultimately hope to make useful research recommendations about PD design, we need to anticipate what factors will (or will not) influence faculty's movement of ideas, practices, and/or interactional norms across PD and classroom contexts. Thus, we broaden our view to take up theoretical elements from Vygotsky and the many scholars who have interpreted and expanded on his work [Vygotsky, 1980, Daniels, 2008, Cole, 1998, Rogoff, 1990, Lave, 1996, Hutchins, 1995].

From a Vygotskian perspective, interactions with others and physical tools can serve as mediators of thought, and should be considered an integral (and inseparable) part of thinking and learning—it is not useful to divorce cognition and action [Vygotsky, 1980, Hutchins, 1995, Lave, 1996]. While individuals' thoughts play a role in their learning, the interactions of people with tools, and people with other people, are the developmental precursors of what will later be internalized by individuals, and thus analyzing these interactions is a valuable—perhaps critical—way to document learning. This perspective suggests that modeling the form and structure of distributed cognitive processes can give us insight into what faculty will take away from these interactions: it is not necessary for our research to also document individual faculty's thinking after these interactions in order to be valuable and credible, because the interactions that happen in the moment are a part of the learning process and deserve significant attention in their own right. We want to know how faculty are influenced by their surroundings and the activities they are embedded in because this is a critical part of their learning, and because PD leaders play a central role in shaping these interactions and therefore could use this kind of research to inform productive shifts in PD design. During PD, faculty's ways of reasoning are situated in workshop sessions, and thus depend on what ideas they are encouraged to contribute, how workshop activities are launched, what tasks faculty are asked to engage with, how faculty interact with each other, and what workshop leaders say or do to facilitate faculty's interactions. Returning to the previous example we presented where faculty act as pseudo-students, we can broaden our view to centrally include more aspects of the workshop session in our analysis. For instance, some science tasks may be more or less conducive to writing down correct answers without understanding, faculty's interactions with a workshop leader or their peers could support or disrupt school-like norms, and these interactions with people and tasks can be thought of as emergent experiences that are part of faculty's learning about teaching.

Another common definition of "resources" in education research, specifically research focused on teacher education, can allow us to expand our use of the term "resources" to be more aligned with the Vygotskian tradition and help us to identify potential bridges between PD and classroom contexts. Cohen et al. [2003]'s highly-valued research establishes a wide range of teacher "resources" that can be categorized as conventional, personal, and environmental or social. For faculty, conventional resources would include physical resources like textbooks or other curricular materials, classroom space, and class size; personal resources would include both faculty's knowledge and interests, as we have been discussing, as well as students' knowledge and interests; and environmental or social resources would include local academic norms and social support from peers or administrators. All of these "resources" could importantly shape faculty's instruction. These broadly-defined resources might launch or sustain various ways of faculty thinking, and could either support or discourage faculty's use of research-based instructional strategies and principles. For instance, students' expectations and questions, social pressures from colleagues, and inherited curricular materials for a particular course might promote research-based instruction, but it is also likely that they would pull faculty towards more traditional instruction [Ball and Cohen, 1999].

Because faculty will likely need to navigate which local resources they draw support from and how to do this, an important role of PD could be to help faculty to relate their PD experiences to their experiences at their home institutions in ways that lend stability to their continued use of progressive, research-based teaching methods. Because we want to explore the extent to which this is occurring in PD, we need to take a theoretical stance that acknowledges the diverse set of resources that faculty could have available to them. From this theoretical perspective, we already note that greater similarity between faculty's classroom contexts and PD workshops will likely generate increased opportunities for faculty to identify local resources that could serve as anchors for improving their instruction. We will start to consider some PD designs that might promote this coherence across PD and classroom contexts in the following section.

As situations arise where it strengthens our analysis to label particular resources, we will more carefully indicate how we are using this term. For now, we simply note that many of the "resources" mentioned here could be relevant in our analysis and that we consider all to be important in considering potential shifts in faculty's instruction.

2.3 Conceptualizing the design of learning environments for faculty

So far, we have made a number of statements about what we might want faculty to become able to do in their classrooms in a broad sense, how one might think about addressing the limitations of typical approaches to instructional change within PAER, and what could constitute a useful approach to conceptualizing faculty learning. In Chapter 1, we stated that in order for the astronomy and physics communities to make the most progress in creating more equitable experiences for our students, faculty need to become partners in instructional change: they need to flexibly adapt RBIS to fit their local circumstances, and they need to be able to fill in gaps in what can be prescribed by education researchers. We suggested that these gaps exist both because there is research left unexplored and because there are fundamental limits to how much we should specify in advance. At the same time, faculty's enculturation into traditional instruction likely makes it difficult for them to envision how research-based recommendations can be enacted in their classrooms, or to fully embrace unfamiliar strategies quickly [Lortie and Clement, 1975]. Because of this, we echoed the sentiment that faculty need training and support, and would likely benefit from increased engagement with the education research community over long timescales.

In Section 2.1 of this chapter, we outlined the landscape of approaches to instructional change in undergraduate STEM education, particularly physics and astronomy, based on Henderson et al. [2011]'s empirically-based framework. We noted that most instructional change efforts in PAER have been prescribed and aimed at individual faculty, such as workshops focusing on the dissemination of RBIS. These efforts have led to the positive but limited outcomes that others have documented: faculty gain awareness and motivation to try RBIS, but often do not persist in using these strategies in the long term and may modify RBIS in ways that make them less likely to foster the positive student outcomes the designers observed. We suggested that when it comes to faculty PD workshops, deviations towards more emergent goals might improve workshop outcomes. Holistically, we suggested that

PD workshops might be better conceptualized as a piece in faculty's learning over long timescales, and PD leaders could work towards building faculty's capacity to engage with education research ideas and research-based pedagogies in productive ways when they return to their home institutions.

In Section 2.2 of this chapter, we outlined how one could conceptualize faculty's learning in useful ways that move beyond a purely cognitive and misconceptions-oriented view. Specifically, we argued that a useful orientation to faculty learning assumes that all faculty have a wide variety of potentially useful resources that they can draw on, that what faculty will think at a particular moment depends on their perceptions of the context they are embedded in, and that interactions with people and tools are a critical piece of what constitutes faculty's thinking and learning.

While these statements provide a foundation for how we might think about PD design, we have yet to say anything specific about how faculty might need to learn at PD workshops such that they leave better prepared to engage with and wisely adapt education research strategies at their home institutions, or about how one might design PD environments that foster this kind of learning. In this section, we will start to consider these more practical aspects of PD design. We first introduce language from Sandoval [2014] that we will draw on to distinguish between the various elements of PD workshop design: embodiment, mediating processes, and outcomes. We consider what we can learn about these design elements from existing literature by summarizing K-12 PD goals that align with and substantiate our vision for faculty's learning, PD activities that have been proposed as viable ways to achieve these goals, and the extent to which evidence exists to suggest that these activities can lead to the desired outcomes. We then return to Sandoval [2014]'s work to argue for the importance of hypothesizing about how design is linked to the interactions we have positioned as central to faculty's learning, and about how faculty's interactions in PD are linked to particular outcomes. We demonstrate that other researchers have followed similar logic using two examples from the K-12 literature. Finally, we turn to faculty PD workshops and unpack two examples in physics and astronomy using this model, and use these examples to introduce and motivate the analysis we will pursue in this thesis.

2.3.1 Conceptualizing design

Here, we will take a moment to introduce Sandoval [2014]'s work, which provides a useful way to delineate different elements of PD design. In particular, he defines three distinct components of educational design: embodiment, mediating processes, and outcomes. For us, *embodiment* would describe the ways that PD is structured—what we might colloquially refer to as the design—which would include the specific tasks or activities that participants are asked to engage with, the ways that interactions are structured (e.g., how and to what extent participants are expected to interact with each other and the workshop leader), and how the workshop leaders' facilitation choices might be characterized (e.g., the extent to which workshop leaders steer conversations in a predetermined direction). *Mediating processes* would describe what participants think, say, and do as they interact with other people and tools during PD. Finally, *outcomes* would describe what changes for participants as a result of participating in PD. When we discuss design here, we largely do so from the perspective of imagining how one might think about creating PD experiences, and therefore will first discuss goals rather than outcomes, then turn to the extent to which there is empirical evidence that these outcomes have emerged as planned. PD leaders' explicit goals help us to infer the intended PD outcomes.

2.3.2 Ambitious PD goals

Several teacher PD goals that have been primarily articulated in the K-12 literature are commensurate with our vision for ideal faculty PD outcomes. While at other points in this thesis, we will consider a wider range of possible goals (e.g., increased awareness of RBIS, increased motivation to try RBIS, etc.) in order to capture the breadth of potentially worthwhile activities that occur in faculty PD workshops, here, we explicitly focus only on ambitious PD goals that we think deserve increased attention within the PAER community.

One ambitious and emergent goal is *to improve instructors' abilities to notice student behaviors and ideas during class, including how students respond to instructional moves or work with their peers during class* [van Es and Sherin, 2002, Sherin and van Es, 2008, van Es and Sherin, 2008, 2010, Scherr and Goertzen, 2015, Robertson, 2015]. Researchers have described noticing as a component of instructors developing "professional vision"[Goodwin, 1994], or ways of seeing the classroom in more "expert" ways.

Another related goal is *to improve instructors' abilities to elicit and respond to students' ideas during class* [Robertson, 2015, Brodie, 2011, Coffey et al., 2011, Richards, 2013, Hammer et al., 2012]. Many prescribed RBIS provide a strong start—for example, we previously described PI and other RBIS as structured mechanisms that help instructors to do this—but some aspects of interacting with students in ways that align with these ideals likely require instructors to have more skill and flexibility than RBIS developers can fully prescribe [Ball and Cohen, 1999, Hawley and Valli, 2000]. In particular, it may be desirable for instructors to adapt their instruction in-the-moment based on emergent student talk, which requires them to be responsive to students' ideas. Outside the classroom, we might want instructors *to become better able to use student responses to instruction, including students' emergent disciplinary ideas, to plan next steps* [Robertson, 2015], which also relies on their ability to notice what occurs in the classroom.

In this longer-timescale reflection, we might want instructors *to become better able to use education research principles to guide their planning by considering how these principles apply to the concrete examples of practice that they experience, and to hypothesize about how alternative instructional choices could affect students' behaviors* [Horn and Little, 2010, Aubusson et al., 2010, Morrell and Schepige, 2012]. This planning could involve thinking about maintaining or shifting pedagogical strategies (e.g., hypothesizing about how questioning and responding to students in different ways might be consequential), or it could involve selecting, refining, or creating disciplinary tasks to be better suited to students' emergent strengths and needs (e.g., hypothesizing about what disciplinary tasks will be appropriate for their students at a particular moment) [Ball and Cohen, 1999].

All of these could be thought of as ambitious goals focused on cultivating pedagogical practices for individual instructors. We agree that helping individual instructors to improve their own instruction through each of these mechanisms is laudable by itself. However, at the same time, we could think of PD as a way to improve instructors' abilities to initiate or sustain future, productive pedagogical conversations by engaging other local educators in the reflective practices described above. If we strongly take up the premise that faculty's learning occurs through interactions with others, and that learning about instruction is a long-timescale process, then a reasonable goal of short PD experiences might be to enable faculty's learning to continue by modeling productive conversations about instruction and thus motivating and enabling faculty to pursue these kinds of conversations in the future. In this way, we can see the goal of developing instructors' "professional vision" as extending to developing instructors' collective ways of learning from emergent instructional situations and "seeing" the classroom together [Goodwin, 1994]. Like [Ball and Cohen, 1999], we think PD provides an opportunity to work towards shifting the discourse within the instructional community.

2.3.3 Correlating K-12 PD design structures to documented outcomes

Many researchers have suggested general kinds of PD activities or designs that could support the ambitious goals we have just described. While we consider both

K-12 and faculty professional development literature throughout this section, we note that there may be differences in how productive teacher interactions can be structured and supported that are contingent on differences in the incoming knowledge and preparation of faculty versus K-12 teachers; for instance, faculty likely have more extensive content knowledge but less pedagogical training than K-12 teachers, and this would likely lead to differences in how the same PD activities would play out. However, we argue that there are also strong underlying similarities between the kinds of teacher interactions that PD leaders might want to support to foster the outcomes outlined above which span both faculty and K-12 teacher preparation settings, as well as similarities the basic design features that could support these interactions. In particular, one relevant takeaway that is pervasive across PD design recommendations is that PD experiences should be centered on practice, i.e., PD activities should be as similar as possible to real, specific situations that teachers regularly encounter [Ball and Cohen, 1999, Loucks-Horsley et al., 2009, Darling-Hammond et al., 2009, Hawley and Valli, 2000, Garet et al., 2001, Desimone et al., 2002]. This could happen in a variety of ways, including watching and discussing classroom video, experiencing and discussing mock examples of classroom practice (acting as students or as educators), analyzing written case studies or teacher notes, unpacking or designing disciplinary tasks for students, examining real student work, and planning next steps for how to incorporate new curricular materials into their instruction [Ball and Cohen, 1999, Darling-Hammond et al., 2009, Loucks-Horsley et al., 2009]. Some have also suggested that there might be affordances to having PD participants contribute classroom video or student work from their own classrooms for use in these PD activities [Darling-Hammond et al., 2009, Ball and Cohen, 1999]. It is clear to us that the way PD is structured matters for what instructors will learn, and it is highly plausible to us that the instructor abilities we listed could be developed and embedded within these activities.

Although we see significant promise in these proposed activities, and some K-12 studies broadly show positive correlations between these kinds of PD design choices and desirable teacher outcomes, less work has been done to develop detailed, mechanistic explanations for why and how these activities lead to improvements in teacher practice. For example, a large, statistical study by Garet et al. [2001] substantiate others' claims that "active learning," a focus on disciplinary content knowledge, and coherence with teachers' local situations comprise three key features of K-12 PD by demonstrating correlations between these features and enhanced teacher knowledge and skills. Their use of the term "active learning" encompasses essentially all of the PD activities described in the previous paragraph. Similarly, [Darling-Hammond et al., 2009] review existing K-12 PD literature and summarize proposed links between PD activities and teacher outcomes: for example, various studies have provided evidence that shifts in teacher practice are correlated with teachers seeing instructional practices modeled in PD settings; defining learning goals and potential student difficulties surrounding disciplinary content; observing and giving feedback on their peers' instruction and vice versa; and analyzing video of their own instruction. While these studies provide initial evidence that these activities can be effective, many details of how these PD experiences were implemented, how teachers interacted with other people and tools during PD (Sandoval [2014]'s "mediating processes"), what changed about instructors' knowledge or skills, and, perhaps most critically, associated theoretical claims that could illuminate what specific aspects of these PD experience seemed to be consequential for particular outcomes, are largely left unexplored. In other words, many of these studies seem to lack detailed models of instructors' learning within PD settings that could guide future PD leaders in replicating past results or making informed adaptations.

2.3.4 Conjecture mapping

In order to explain how one might strive to avoid these potential roadblocks or limitations in PD research, we revisit Sandoval [2014]'s work from the start of this section. In addition to providing us with a language to describe the components of learning environments, Sandoval [2014] strongly supports and shapes our claims about what is missing from some of the studies above. He defines an analytical framework that he calls *conjecture mapping*, which allows researchers to articulate and test hypotheses about how learning is supported in a given context (as is central in design-based research [The Design-Based Research Collective, 2003, Cobb et al., 2003]). In particular, he argues that it is critical to make conjectures about the ways that the embodiment of a particular design is linked to mediating processes, which he calls *design conjectures*, and the ways that these mediating processes are linked to outcomes, which he calls *theoretical conjectures*. In line with Sandoval [2014], we argue that building these models to describe PD efforts is a critical step towards understanding how they support instructors' learning and what modifications could make them more successful. We also agree with Sandoval [2014] that from a practical perspective, foregrounding certain pieces of this model empirically may be necessary to support the development of detailed and robust claims. This does not preclude the utility of articulating thoughtful hypotheses that define and connect any of these elements, even with limited empirical evidence. We also note that Sandoval [2014]'s model accurately captures the kinds of choices and assumptions that PD designers must make when initially planning and implementing workshops, even if they do not articulate these assumptions clearly, and therefore it is valuable for researchers to highlight and discuss these hypothetical causal links further.

Although we have been somewhat critical of others' approaches to researching PD, some notable exceptions exist where K-12 teacher PD has been carefully modeled in ways that we consider to be consistent with Sandoval [2014]'s suggestions. Horn and Little [2010], Horn [2010], Horn and Kane [2015], van Es and Sherin [2002], Sherin and van Es [2008], van Es and Sherin [2008, 2010], Aubusson et al. [2010], Morrell and Schepige [2012], Harlow [2010], and Ball and Cohen [1999] each articulate models of how teacher learning occurs in PD alongside detailed empirical evidence, and could be described as building conjectures between design structures and mediating processes, and/or between mediating processes and teacher outcomes. For brevity, we will summarize two of these studies—Ball and Cohen [1999] and Horn and Little [2010]—that relate to our analytical focus in Chapter 5. As we summarize these studies, we will situate them within Sandoval [2014]'s conjecture mapping framework.

First, two empirically-grounded PD vignettes from Ball and Cohen [1999] include each of the pieces that Sandoval [2014] laid out, and tell theory-driven stories that highlight the potential value in design characteristics that are shared across these examples. In particular, Ball and Cohen [1999] chose PD examples that have both similarities and differences in their embodiment, but have similar mediating processes and outcomes. In both examples, teachers form workgroups, group members pursue shared and/or similar goals related to mathematics instruction, and their work is grounded in practice in some way. Variations occur in teachers' initial goals and the tasks that they engage with: in the first case, teachers aim to improve their students' performance on a state assessment, a task that has immediate and real consequences, and choose to examine their own students' work from the previous year's test; in the second case, teachers aim to improve their ability to notice, support, and assess students' writing about mathematics, a task that is more driven by curiosity and general interest in improving instruction, and choose to examine students' work from other instructors' classes. When it comes to mediating processes, Ball and Cohen [1999] claim that in both cases, teachers' varied and sometimes conflicting ideas are voiced and debated in these workgroups, which provides rich opportunities for individual teachers to learn. Because similar mediating processes emerge, it is not surprising that similar outcomes arise as well: teachers in both groups find the work enjoyable and productive towards helping them to understanding student thinking, and both teacher groups choose to continue their work together over longer timescales than originally planned. Ball and Cohen [1999] thus implicitly outline a design conjecture that states: if teachers examine student work in light of an authentic teacher-driven question, then teachers will voice and debate varied ideas about pedagogy; and a theoretical conjecture that states: if these varied and conflicting ideas are voiced and debated, then teachers will persist in their participation, enjoy their collaborative work, and deepen understanding of student thinking. These conjectures could then be iterated on by other PD researchers or PD leaders.

As a second example from the literature, Horn and Little [2010] undertake a comprehensive study that illustrates how potentially consequential PD design characteristics can be identified through examining contrasting cases. The authors study two teacher workgroups, and find that unlike in the Ball and Cohen [1999] cases, one group made significantly more progress towards their shared goals (largely based on their conversations) and thus could be characterized as more productive than the other. Based on observed differences in conversational routines, along with additional information gathered to contextualize these observations, the authors hypothesize that three structural factors ("embodiment"), present in one group and not the other, seem to support useful interactions ("mediating processes") and thus create rich learning opportunities for teachers in that group.

Specifically, the three supporting, structural factors they identify are:

(1) teachers' abilities and tendencies to use shared educational concepts, principles, and language to structure their conversations and guide interpretations of instructional challenges, which were developed during the teachers' prior, shared PD experiences;

(2) access to a common, coherent set of curricular materials, as opposed to incomplete materials that lead individual teachers to make last-minute additions or

changes;

(3) leaders among the group who work to establish and maintain productive group routines by posing questions and eliciting ideas from others, treating both student and teacher learning as relevant topics, and legitimizing specific instructional challenges as worthy of the group's attention instead of only normalizing them.

Because Horn and Little [2010] identify these salient differences in PD embodiment based on the conversational routines (i.e., mediating processes) they observe, and then argue for how these conversational routines are consequential for teacher outcomes, we consider this a strong example of conjecture mapping that focuses on developing robust design conjectures. We can summarize Horn and Little [2010]'s conjectures as: if teachers share a principled pedagogical toolkit, have concrete and coherent curricular materials on hand, and work with leaders who enforce group norms, then productive group routines will emerge; and if productive group routines emerge, then rich opportunities for teacher growth will arise. Horn and Little [2010] also hypothesize that if these embodiment structures are lacking, productive group routines will not arise; and if productive group routines to do not arise, learning opportunities will be more limited.

2.3.5 Modeling faculty PD workshops

Now that we have reviewed others' approaches to designing and researching PD in K-12 settings and laid out an analytical framework that guides how we will conceptualize PD design ourselves, we are well-positioned to look more narrowly at modeling faculty PD workshops. As we have alluded to previously, it is rare for others to have done modeling work for faculty PD, and this shortcoming has been a key driver of this thesis work. However, there are two notable cases where models have been proposed by other education researchers and/or the PD designers themselves [Prather et al., 2009, Chasteen et al., 2011]. In these cases, we can use external representations of their PD design (as embodied in handouts, worksheets, YouTube videos, journal articles, PowerPoint slides, etc.) as well as our field note observations of workshop implementation to illustrate how Sandoval [2014]'s framework may lend insight into the potential outcomes. Thus, we choose to launch our examination of faculty PD workshop design by illustrating these two cases as situated within Sandoval [2014]'s framework.

The first case we will consider is Eric Mazur's talk entitled "Confessions of a Converted Lecturer." As we mentioned briefly in Section 2.1, this talk is a keystone example of the "Disseminating Curriculum and Pedagogy" change strategy within PER: a YouTube video of this talk has been viewed over 100,000 times, and Mazur has traveled to many locations around the world (on commission) to repeat it in person. In his talk (the embodiment of this design), Mazur tells a personal narrative of his past teaching struggles—his prior hubris surrounding his ability to teach effectively through lecture, his surprise at a dearth of student conceptual learning as measured by a PER assessment instrument following his lectures, and his emotional reactions to both this lack of learning; his personal discovery that students may be more able to learn from their peers when they do not understand his physics lectures; and his later teaching successes, including significantly improved student conceptual learning and student buy-in to his revised approach.

This storytelling is done in service of introducing a particular RBIS, Peer Instruction (PI) [Mazur, 1997], that Mazur enacted as a result of these personal struggles and discoveries. When it comes to "mediating processes", Chasteen [2011] has hypothesized that Mazur's talk has so much pull because faculty see him as credible, trustworthy, and similar to them as they listen to this lecture [Olson, 2009]. As she suggests, these perceptions seem tied to Mazur's position as a Harvard professor, his initial skepticism and lack-of-use of education research, and his reflection on teaching struggles that many instructors also face, compounded with the light-hearted way in which he recounts his past experiences. Because Mazur typically elicits these perceptions from faculty and weaves his experiences together into a compelling story, faculty listen, pay close attention, and are entertained, generally "buying-in" to his message. When it comes to "outcomes", this experience may lead many faculty, in Mazur's language, to get "all fired up" and try PI in their own classes. Significant evidence supports this finding: PI is the most popular RBIS in physics [Henderson and Dancy, 2009, Dancy et al., 2016], and attending the NFW, where Mazur usually presents, is the strongest predictor of faculty being aware of and trying out RBIS in their classes [Henderson et al., 2012]. But, as we also described in Section 2.1, faculty often do not persist in using these RBIS, and it has been hypothesized that faculty who try PI as a result of listening to Mazur's talk face unanticipated implementation challenges [Henderson et al., 2012, Turpen et al., 2016, Dancy et al., 2016, Chasteen, 2011].

We can summarize these conjectures as: if a prestigious and highly respected physicist tells a compelling personal narrative of educational enlightenment, then faculty will actively listen and relate to this story; and if faculty actively listen and make personal meaning of this account, then they will buy-in to the pedagogical change, be interested in taking up new strategies, and try these new strategies in their classrooms.

Prather et al. [2009] lay out the logic for an alternative PD workshop approach, which they call a "situated apprenticeship" model, that they argue has the potential to address the shortcoming of faculty not understanding RBIS well enough to implement them successfully in their classrooms. To illustrate this model, Prather et al. [2009] describe their own workshop, the Center for Astronomy Education (CAE) Tier I Teaching Excellence Workshop, which has also been enacted in an abridged form at the NFW. As our second case, we will consider the subset of their PD model that focuses on teaching instructors how to implement what is essentially the same strategy as PI "live" in the classroom. (For simplicity, we omit discussion of teaching instructors to create questions for this strategy, as well as other, more tangentially relevant elements of their workshop.) Here, PI is labeled as Think-Pair-Share (TPS) and associated with a more detailed (and sometimes slightly different) set of prescribed steps than the ones Mazur outlines; we adopt Prather et al. [2009]'s label to indicate this second distinction.

In the embodiment of this workshop design, participants first experience TPS as pseudo-students while a workshop leader acts as the instructor, then take turns trying out the strategy as instructors themselves while their peers again act as pseudo-students and the workshop leader observes. The workshop leader articulates a justification for each step in TPS when they model it for participants initially, and/or elicits justifications from participants. When participants act as the instructor, their peers are instructed to act as critical friends who "pause" the mock implementation (by saying "pause") when they notice deviations from the prescribed script and provide feedback that redirects the "instructor." The workshop leader works to establish these norms by saying "pause" themselves early on and prompting participants to fill in their reasoning. As anticipated mediating processes, participants experience what it feels like to enact TPS and be students in this pseudo-classroom setting, listen to justifications for the prescribed TPS steps, notice potential variations in TPS implementation with increasing regularity, and potentially "pause" their peers and articulate justifications for the prescribed steps themselves. Prather et al. [2009] hypothesize that the outcome of these mediating processes will be that participants will gain deep knowledge of what direct adoption of TPS looks like and of its potential pedagogical affordances, and could choose to implement TPS in their classrooms with a high degree of fidelity and thus a high likelihood of success.

Thus, in the language of conjecture mapping, Prather et al. [2009] present the hypotheses that if faculty experience TPS as pseudo-students and mock instructors, and experience giving critical feedback on mock implementation, then faculty will engage in analyzing TPS from multiple vantage points; and if faculty engage in critical analysis of TPS, then they will be able to successfully adopt TPS in their classrooms and justify their implementation.

In comparing these two PD models, we agree with Prather et al. [2009] that in their model, faculty are likely to gain a significantly deeper understanding of how to adopt PI/TPS, which might better prepare them for challenges they could face in the classroom and thus help them to persist in using this strategy. We also note that Prather et al. [2009]'s model follows the PD design recommendation of striving to make PD experiences similar to instructors' local experiences by simulating classroom instruction. We find this design characteristic to be theoretically salient as well: greater similarity between faculty's classrooms and PD contexts could make it easier for faculty to learn to draw on resources that would be readily available to them in their local contexts, such as their students' ideas or the steps in the prescribed TPS implementation.

We can see significant complexities and potential challenges that could arise in the precise embodiment and resulting mediating processes of these and other faculty PD designs. In particular, we have already stated that a focus on adoption might contrast with faculty's inclination to modify and adapt RBIS [Henderson and Dancy, 2008]. It is clear to us that teaching faculty how to adopt RBIS is often the central goal of faculty PD, but it is not clear whether and to what extent workshop leaders create space for faculty to genuinely consider alternative suggestions and weigh both potential affordances and drawbacks of prescribed strategies. Like Ball and Cohen [1999], we anticipate that infusing PD implementation with debate and critical analysis of possible instructional choices instead of only giving prescribed answers will be consequential for some of the ambitious goals that we orient to, such as supporting faculty in learning to flexibly adapt to match their students' emergent behaviors. These potential complexities warrant empirical investigation of workshop implementation and associated faculty engagement.

In our work, we aim to interrogate what actually occurs within PD workshops, and to further support the envisioning of PD that goes beyond increasing faculty's awareness and motivation to try RBIS. In part, this will involve providing access to a detailed picture of how workshop sessions are embodied, so that PD leaders and researchers will have an evidence-based starting point for making claims about what mediating processes and outcomes are plausible. To this end, we have created a tool that allows others to document the structure of other workshop sessions, and we demonstrate how one could build plausibility arguments for PD outcomes based on this documentation of other designs. We also closely investigate the embodiment and resulting mediating processes of sessions that seem like they could plausibly support ambitious outcomes. Through this detailed qualitative analysis, we will begin to investigate the kinds of complexities we mentioned in the previous paragraph. This qualitative work will also provide workshop leaders with concrete examples of facilitation moves they might want to enact or avoid in order to promote interactions that seem to support faculty's learning in certain ways within these settings.

2.4 Thesis overview

We have now reviewed the empirical, theoretical, and analytical frameworks we that we use to guide this thesis work and situate it within existing literature. In this final section of Chapter 2, we outline the remaining chapters that comprise the body of this work. In Chapter 3, we describe the data we collected during this project, which directly enables our analysis and informs our thinking more broadly. In Chapter 4, we describe the development of a new instrument, the Real-time Professional Development Observation Tool (R-PDOT), which allows an observer to document workshop design (or embodiment, in Sandoval [2014]'s language) in ways that draw attention to the interactivity and prescriptiveness of these sessions and capture the kinds of activities we might want or expect faculty to engage in. We demonstrate the potential utility of this tool for PD leaders by showing R-PDOT data from three example sessions from the NFW, our primary research site, and illustrate how this data allows us to build grounded hypotheses about what faculty interactions and outcomes might emerge as a result of attending each session. In Chapter 5, we pursue qualitative analysis of several episodes within the NFW that contained significant amounts of faculty talk based on initial data collected with the R-PDOT, and, not coincidentally, where we are able to find compelling evidence of faculty's thinking and learning based on their interactions. In particular, we identify markers of faculty orienting to the session content in traditional school-like ways versus taking up a focus on making sense of pedagogy, and analyze how the workshop leader's facilitation moves seem to cue up, support, or disrupt these different orientations. In Chapter 6, we turn to a different setting and change approach, and overview the results of and process behind the local transformation work that we undertook surrounding an introductory astronomy course sequence for majors. We conclude with recommendations for PD leaders, researchers, and other proponents of change in undergraduate STEM education in Chapter 7.

Chapter 3: Data collection

Video recordings from the Physics and Astronomy New Faculty Workshop (NFW) are the primary data source for Chapters 4 and 5, which comprise the bulk of this thesis work. The NFW is highly valued within our community: it is sponsored by the three professional societies—the American Association of Physics Teachers, the American Physical Society, and the American Astronomical Society; it has been continuously funded by multiple consecutive grants from the National Science Foundation over the past 20 years (since 1996); and it has been attended by between 25-50% of new tenure-track hires in physics during that time [Henderson, 2008, Council of Scientific Society Presidents, 2012]. Because of this, it presents an ideal opportunity to interrogate standard approaches to PD workshop design for undergraduate physics education that will be familiar to many. This 4-day workshop is currently run twice each year, in June and November, and has been attended by 50-70 faculty members during each implementation in the times we have observed it. It is split into many 45-90 minute sessions, most focused on teaching faculty about specific RBIS, and each session is facilitated by one (or more) of the 15-20 invited presenters, who are often the original developers of the focal RBIS. Each of these workshop leaders has considerable freedom in how they design and lead their sessions; thus, observing multiple sessions allows us to see a cross-section of different PD approaches. More recently, new parallel sessions have been added to mirror the Center for Astronomy Education (CAE) Tier I Teaching Excellence workshop, which has allowed us to study how Prather et al. [2009]'s PD model that we discussed in Section 2.3.5 plays out when it is enacted under different constraints and facilitated by different workshop leaders who take different approaches to enacting the "same" design.

The NFW is held at the American Institute of Physics in College Park, Maryland, which is local for our research team, just down the road from the University of Maryland. This has enabled me to attend and at minimum take field notes during every iteration of the NFW from June 2013, the start of this project, through November 2015. We captured video of a majority of the instruction-focused sessions across three iterations of the NFW, which included approximately 2 dozen uniquely designed sessions and several instances where we captured the same workshop leader enacting highly similar session designs multiple times. In each instance, we capture the workshop leader and faculty participants from the back of the room. When faculty worked in small groups for extended periods of time (more than about a minute), we mic-ed a focal group and captured their discussion on video, either using a second camera or the primary camera (depending on the setup of the room). As we developed our workshop observation tool (the R-PDOT) as discussed in Chapter 4, we both tested out our codes in real-time during workshop sessions and used videos that captured the workshop leader and participants together to iteratively refine the tool and establish inter-rater reliability. In Chapter 5, we analyze video recordings of extended small group work that contained compelling evidence of faculty's thinking and learning, as well as video of discussion and interactions among all faculty participants and a workshop leader (when faculty were divided into 3-4 smaller groups of 15-25 participants for "breakout" sessions). Because each of these body chapters draws on video data in different ways and employs highly different analytical methods, we will wait to describe and justify our data selection further within these chapters as it becomes relevant to the analysis that immediately follows.

While these video recordings were the central focus of our formal analysis, we collected a variety of other data that informed our thinking. At the start of this project, we attended and collected data at two iterations of the CAE workshop as we piloted and refined our data collection strategies and started to conceptualize the R-PDOT. We have conducted about a dozen formal one-on-one interviews and had many informal conversations with workshop presenters at the CAE workshop and the NFW, including several interviews that helped us to understand others' interpretations of R-PDOT data and refine it to be more accessible to our target audience in its late stages (as discussed in Chapter 4). We have also taken steps to better understand these workshops from the perspective of participants: we interviewed roughly three dozen faculty about their teaching and their workshop experiences; we took extensive field notes during workshop sessions that focus on capturing faculty's engagement (e.g., to what extent they seemed alert and engaged, and what kinds of questions they asked) and following conversations with participants during breaks; and we coordinated our efforts with the NSF external evaluators of the NFW, initially Dr. Charles Henderson and currently Dr. Stephanie Chasteen, to shape the pre-/post- workshop survey for NFW in ways that aligned with our initial interests. While we have not systematically analyzed these data sources in depth, the knowledge we gained informed the development of the R-PDOT and has helped to shape the kinds of claims we will make throughout this thesis. We note that while we draw attention to a wide variety of possible PD workshop components when defining the R-PDOT codes in a broad sense, we do not fully elaborate on the variety of promising session designs that we have observed, and we recognize that other promising approaches to PD design are likely being enacted in settings that we have not observed. In particular, while NFW sessions do represent of a variety of approaches to workshop design within PAER because they are led by a variety of facilitators, we also recognize that many of the NFW facilitators have run their own, longer workshops where they have significantly more autonomy and flexibility to engage faculty in learning about teaching in other ways. While we consider this investigation of the NFW to be a logical first step, we consider research into faculty's engagement within these other settings to be a worthwhile future pursuit.

In Chapter 6, we use more canonical PAER methods to document and analyze the successes and limitations of our local course transformation efforts, such as research-based pre-/post- class assessment data using the Light and Spectroscopy Concept Inventory (LSCI) [Bardar et al., 2005] and the Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS) [Zwickl et al., 2012]), student responses to evaluations, data describing student and instructor behaviors during class collected with the Classroom Observation Protocol for Undergraduate STEM (COPUS) [Smith et al., 2013], written reflections on instruction, and field notes taken by our team members during class. We also use email exchanges, past team meeting agendas, and drafts of course materials to substantiate our recollection of the change process.

Chapter 4: Assessing the interactivity and prescriptiveness of faculty professional development workshops: The Real-Time Professional Development Observation Tool (R-PDOT)

We note that this chapter has been submitted for publication in the Physical Review-Physics Education Research and is included verbatim. As a result, some of the introductory content is repeated from earlier chapters.

4.1 Introduction

There is a general consensus among national policy makers and education researchers that undergraduate STEM instruction can be improved through closer alignment between faculty's teaching and education research principles and findings [Singer et al., 2012, President's Council of Advisors on Science and Technology, 2012]. For many faculty, the easiest and most efficient path towards this alignment is to adopt or adapt existing research-based instructional strategies (RBIS) in their classrooms. Consistent with this idea, many faculty professional development (PD) efforts—particularly efforts led by the discipline-based education research community—have focused on disseminating RBIS [Henderson et al., 2011] and have been successful to an extent, yet we rarely critically examine these efforts to understand what contributes to these successes and how we could improve. For example, prior research has shown that physics faculty are more likely to be aware of and experiment with RBIS in their classrooms after attending the Physics and Astronomy New Faculty Workshop, but that many faculty find it difficult to persist in using these strategies over time [Henderson, 2008, Henderson et al., 2012]. This lack of persistence could indicate both a need for long-term support and ways in which existing short-term efforts do not address faculty's needs and concerns. Scientific society leaders in biology, chemistry, mathematics, geoscience, and engineering all sponsor well-attended, discipline-specific faculty workshops akin to the New Faculty Workshop [Council of Scientific Society Presidents, 2012], and encounter similar limitations and successes [Ebert-May et al., 2011, Borrego et al., 2010, Macdonald et al., 2005].

No in situ research has been conducted to directly investigate what occurs during faculty PD workshops like these large national workshops, or—to our knowledge—local, small, and/or non-discipline-specific workshops. Because of this, PD leaders have neither concrete evidence of what specific PD experiences could lead to specific outcomes [Sandoval, 2014], nor tools to help them justify and communicate about their design decisions in robust and consistent ways. Moreover, indirect evidence and reports suggest that workshop leaders often rely on lecture to disseminate ideas about teaching with minimal participant contributions, which leads us to think that workshops could be improved through careful consideration of how alternative, more engaging design choices might produce more desirable shifts in faculty's thinking, practice, or participation in ongoing PD [Loucks-Horsley et al., 2009, Desimone et al., 2002, Garet et al., 2001, Wilson, 2013, Valli and Hawley, 2002]. Despite the lack of in situ observation and analysis of workshops, prior research can lend some insight into what could be contributing to the limitations of workshops (beyond time constraints), and what could be done to address this. Looking across 191 published approaches to faculty PD in STEM, Henderson et al. [2011] find that faculty change efforts typically fall into one of four categories: disseminating curriculum and pedagogy, enacting policy, developing reflective teachers, and developing shared vision. These categories are distinguished from each other by whether they target prescriptive or emergent outcomes, and whether they target individuals or groups. Workshops naturally fall into the prescriptive, individually-focused "dissemination" category, and often encourage faculty to adopt strategies as-is. However, many faculty often want or need to modify RBIS to fit their local contexts, which is not well-addressed by a one-way communication model. Compounding this problem, highly prescriptive professional development may make it difficult for faculty to feel that their experiences and insights matter, which could limit their future engagement with education research communities [Henderson and Dancy, 2008]. These shortcomings of the disseminating curriculum and pedagogy approach are well-addressed by the change strategies of developing shared vision or reflective practice. These approaches to PD, which are more focused on emergent outcomes, are likely critical to faculty's thoughtful adaptation and sustained use of RBIS. Although the four change strategies defined by Henderson et al. [2011] tend to be fairly siloed, we think that workshops could support emergent outcomes by engaging faculty in the kinds of reflection around instruction we would want them to engage in when they teach, and by explicitly helping faculty to identify ways to engage with communities of educators at their home institutions.

In order to support the exploration and analysis of various approaches to PD workshop design and implementation, we have developed an observation tool, the Real-time Professional Development Observation Tool (R-PDOT), which can broadly document what faculty experience when they attend PD workshops. In this paper, we discuss the development of and intentions behind the R-PDOT and demonstrate how its output can be used to reflect on workshop design. The structure of our paper is as follows. In Section 4.2 we describe the theoretical framework that underlies this work. In Section 4.3, we overview research on reflective teaching discussions that informs the structure of and PD elements captured by the R-PDOT. In Section 4.4, we define and justify the specific descriptive codes that comprise the R-PDOT. In Section 4.5, we consider the accessibility of the R-PDOT, including our methods for refining and using the R-PDOT codes, establishing inter-rater reliability, and visualizing data. In Section 4.6, we analyze R-PDOT data from three sessions of the Physics and Astronomy New Faculty Workshop in order to demonstrate how the R-PDOT data can provide a basis for hypothesizing what kinds of faculty outcomes these different designs might support. In Section 4.7, we elaborate on the potential implications of trends we noticed by using the R-PDOT in the Physics and Astronomy New Faculty Workshop, and show preliminary evidence that the R-PDOT data can support workshop leaders in noticing similar key features during discussions about workshop design. We conclude in Section 4.8 by considering how this research could enable members of the education research community to engage in critical conversations about workshop design, and how the R-PDOT could enable future researchers to generate and pursue new research questions about PD.

4.2 Theoretical approach

Broadly speaking, we take a Vygotskian stance that different ways of thinking, seeing, and knowing would likely appear first in distributed form, i.e., in interactions with others, and then internalized by individuals (and transformed in the process) [Vygotsky, 1980, Daniels, 2008, Cole, 1998, Rogoff, 1990]. These interactions can provide the opportunity for a diversity of ideas to be contested, compared, and developed in ways that build meaning and allow individuals to begin to take up perspectives that may have been different from their initial ideas. This theoretical approach is broadly applicable to all learners, be it students, K-12 teachers, faculty, or workshop leaders.

It would be difficult to meet many learners—with a diversity of incoming knowledge and experiences—where they are at through a one-size-fits-all communication approach, like lecture. It is well-recognized within the education research community that active engagement is a positive attribute of classroom teaching, as demonstrated by the vast number of studies within undergraduate STEM education linking increased active engagement to outcomes such as higher student conceptual gains and decreased failure rates [Freeman et al., 2014]. The positive impacts of active learning have also been demonstrated within teacher PD, where increased active engagement within PD is strongly correlated with greater improvements in teacher practice [Garet et al., 2001]. Though similar empirical evidence does not yet exist in higher education PD, we assume the benefits of active engagement extend to faculty learning.

We recognize that the construct *active engagement* is too vague to fully inform the delineation of different types of engagement: both researchers and instructors use this term to refer to a wide variety student interactions and behaviors [Dancy and Henderson, 2007], and some of these are may be more or less valuable for learners. From a Vygotskian perspective, the forms of faculty interactions are consequential to what faculty would learn from these experiences. In considering the form of faculty's interactions, we instead rely on research surrounding *participant structures*, namely, the "configurations of interactional rights and responsibilities that arise within particular activities" [O'Connor and Michaels, 1993, Goodwin, 1990]. Whereas researchers often do not distinguish between variations in types of active engagement, well-defined variations within participant structures have been extensively explored within the literature [Jurow, 2005, O'Connor and Michaels, 1993, Knuth and Peressini, 2001, Lotman, 1988, Bakhtin, 1986, Scott et al., 2006]. We revisit this literature when describing the specific form of the R-PDOT in Section 4.4.

4.3 Background literature

We intend the R-PDOT to capture a range of prescriptive to emergent intended faculty PD outcomes. In creating the tool, we recognize that a variety of PD goals exist, from straightforward goals of raising faculty's awareness of what resources and materials exist, to more ambitious goals of improving faculty's abilities to notice student thinking, reflect on their instruction, and engage in future collaborative discussions around instruction. We relate these more ambitious goals to emergent outcomes, in that the final form of faculty's instruction is not predetermined. We consider awareness-level goals to be easy for workshop leaders to achieve through a variety of mechanisms, including lecture, and therefore do not explore this approach in depth in this background section. In contrast, for more ambitious goals that target emergent outcomes, different kinds of faculty engagement are likely needed, and the form of faculty's interactions matter for what they take away from workshop experiences. We use this background section to explore some PD characteristics that could support ambitious outcomes.

Faculty PD studies are arguably the most relevant to understanding faculty's learning about teaching. Because the participant characteristics in these studies mostly closely match the characteristics of our target population, the claims we make that draw directly from this literature are likely the most robust. Faculty's teaching is often constrained by a variety of factors, including departmental organizational structures, college accreditation, and competing expectations of research excellence, which are not at play in K-12 settings. However, there are goals of faculty PD that significantly overlap with K-12 teacher PD goals, and the endeavor of teaching is fundamentally similar in many ways across undergraduate and K-12 teaching. Both faculty and K-12 teachers likely have routines of practice informed by their prior teaching experiences, and therefore would likely face some common struggles in shifting their thinking and practice, such as learning how to listen to and foster students' potentially productive disciplinary ideas instead of expecting to "fix" or replace students' wrong ideas through lecture. In addition, K-12 teacher PD researchers have made strides in areas that faculty PD researchers have not: for instance, in constructing claims about the nature of instructors' conversations about teaching and what PD activities are often linked to shifts in instruction. While future research that examines faculty's interactions directly would strengthen our claims (since differences between faculty and K-12 teacher's incoming knowledge and preparation may necessitate additional or different support structures to produce similarly productive interactions), currently, the synthesis of this range of literature best allows us to elaborate on how and why certain PD outcomes could reasonably be linked to specific workshop activities.

This literature also informs how we think about supporting workshop leaders in improving their PD practice, which is the overarching goal of creating this workshop observation tool. Analogous to how PD aims to support teachers in fostering student learning in new ways, we aim to support workshop leaders in fostering faculty learning in new ways. Because there is no prior research about how workshop leaders develop their ideas about how to design and implement PD, our tool design decisions are often directly influenced by the literature referenced above (without additional comparisons to literature about workshop leaders' learning or interactions). We suggest that there are sufficiently strong similarities between workshop leaders, faculty, and K-12 teachers, both in the ways they could develop new ways of thinking and noticing through structured discussions with others, and in the ways they could act as facilitators in formal, classroom-like settings, to justify drawing from faculty and K-12 teacher PD research when we consider how to support workshop leaders' thinking about workshop design.

Because we consider reflective discussion around examples of instruction to be a central mechanism for generating new ways of thinking and knowing for teachers (in this case, faculty and workshop leaders), we begin by reviewing relevant faculty and K-12 teacher PD literature on this topic in the following section.

4.3.1 Overview: Cultivating reflective discussion

Here, we review faculty and K-12 teacher PD literature that illuminates the potential impacts of reflective discussions about instruction and the supports that seem necessary for these conversations to be productive. As mentioned above, the literature on reflective teacher discussions influences the development of our observation tool in two important ways: (1) it informs specific aspects of effective faculty PD that we choose to capture with the R-PDOT (described in more detail in 4.4); and (2) it informs the overall form of the R-PDOT, as we intend the tool to serve as an effective vehicle for reflective discussion and debate among workshop leaders.

In both of these cases, reflective discussion can be a critical part of improving instructional practice and preparing instructors for future learning. Yet it would be naïve to encourage any discussions about instruction and expect consistently desirable outcomes. Not all reflection is equally valuable; instead, the form of teacher reflection matters for the degree of resultant instructional improvement. It is all too common for teacher reflection to be insufficiently structured and poorly supported [Hammersley-Fletcher and Orsmond, 2005, Chism, 2007]. For example, faculty peer observations of teaching often focus on superficial tips and tricks that are unlikely to improve student outcomes in a significant way even if taken up by faculty [Hammersley-Fletcher and Orsmond, 2004]. These observations often do not challenge faculty's initial notions of what constitutes effective peer review, and lack concrete examples or other information that could help faculty to operationalize overly general guidelines [Chism, 2007]. Similarly, without guidance, pre-service K-12 teachers often struggle to produce written reflections that adequately prepare them to change aspects of their practice, focusing only on what occurred in the classroom but not why it may have occurred or what lessons they might take away from that experience [Aubusson et al., 2010]. All of these minimally useful observations tend to be highly unstructured, where we use the term *structures* to denote both verbal and written scaffolding that can direct teachers to assess certain aspects of instruction or student engagement.

In the other extreme, if support structures are perceived as too evaluative or judgmental, the likelihood of instructional change can decrease. Specifically, rubric-based observation tools (though helpful in some ways) can create significant barriers to promoting thoughtful reflection and encouraging self-motivated improvement due to their evaluative nature. Teacher resistance to this evaluation is justifiable. At the K-12 level, rubrics are often used for formal teacher assessment rather than for reflective self-improvement, and the outcomes contribute to decisions about career advancement or setbacks. Observers' assessments may be less consistent in practice than anticipated by the designers, different rubrics may yield different assessments of teaching quality for the same class, and there is often significant variation in scores from class-to-class for individual teachers, yet high-stakes assessments are sometimes based on only one observation [Kane and Staiger, 2011, Guarino and Stacy, 2012, Amrein-Beardsley and Osborn Popp, 2011]. Even at the college level, where the stakes are typically quite low, the highly critical feedback that is associated with the implementation of these rubrics often shuts down conversations with instructors and makes them less willing to engage in PD efforts instead of sparking productive conversations around teaching [Hora and Ferrare, 2013, Chism, 2007].

4.3.2 Structures for cultivating reflective discussions

In order to identify some key features that can support both critical and transformative reflections on teaching, here we consider prior research that shows evidence of effective support structures. Many prior studies have demonstrated that teachers can improve when guided to reflect in ways that help them to identify meaningful aspects of their practice to change or maintain [Amrein-Beardsley and Osborn Popp, 2011, Gallos et al., 2005, Morrell and Schepige, 2012, Aubusson et al., 2010, MacIsaac et al., 2001, Hampton and Reiser, 2004, McShannon et al., 2006, Hativa, 1995, Piccinin et al., 2006], and these reflective PD efforts have common elements that motivate the design of our observation tool.

In particular, we find that Aubusson et al. [2010] provide useful terminology that allows us to articulate what kinds of scaffolding can encourage reflection that leads to productive changes to instruction. The authors investigate how pre-service teachers learn to become reflective and what supports them in doing so, and conclude that a focus on *contextual anchors*—specific teacher practices that are observed first-hand (as the instructor) or by watching others—was critical for making the pre-service teacher's reflections generative for their future instruction. They also posit that a focus on *conceptual anchors*—direct connections to education research theory—may be equally important in strengthening teacher reflection. In other words, Aubusson et al. [2010] show that when teachers are able to leverage both observed student behaviors and prior education research when reflecting on instruction, they are able to take next steps that improve their teaching and their students' outcomes.

Other studies in K-12 teacher PD corroborate Aubusson et al. [2010]'s claims. Horn and Little [2010] find that teacher conversations that routinely focus on specific problems encountered in teaching and connect these specific examples to general principles of teaching and learning are more likely to generate viable solutions than conversations that do not. van Es and Sherin [2002, 2008] take a similar stance by defining the ability to connect teaching events to general principles as one of the three key features of teacher *noticing*, and subsequently show that teachers improve their practice as they become more adept at noticing [Sherin and van Es, 2008, van

Es and Sherin, 2010]. In addition to the similarities in what individual reflective practices were developed across these studies, we also note here that all of these studies centrally involve discussion with other instructors, and consider this to be an important piece of how teachers' abilities develop over time.

We can use these ideas to see connections between other studies that involve similar PD elements but do not provide such detailed, mechanistic explanations for their successes. For one, many reflective PD efforts are facilitated by knowledgeable coaches who offer feedback to instructors, and they attribute much of the positive shifts in faculty's attitudes, instruction, and/or student outcomes to this facilitation [Hampton and Reiser, 2004, Hativa, 1995, Piccinin et al., 2006, McShannon et al., 2006, Gallos et al., 2005]. Although PD researchers are often not explicit about what is discussed during instructor-facilitator consultations, it is highly likely that the facilitators, who are sometimes the researchers themselves, are making connections between instructors' current practices and education research theory, and using these connections to centrally inform how they guide instructors to reflect. Moreover, an argument for the importance of finding contextual anchors is consistent with an argument for the importance of teachers and PD facilitators gaining shared knowledge of classroom events, which occurs in many reflective PD efforts [Piccinin et al., 2006, Gallos et al., 2005, McShannon et al., 2006, Hativa, 1995]. In other words, when facilitators learn about how instruction plays out by conducting classroom observations, they are better able to help faculty identify both contextual and conceptual anchors, which together can be used to inform highly productive changes to faculty's instruction.

Structured observation tools or protocols can further enable productive discussions around instruction by providing a guide for the observer as they decide what written and verbal feedback to supply, and a way for the teacher being observed to independently recall key aspects of what occurred in their class based on what is written down, rated, and/or tallied. Rubric-based observation protocols set evaluation criteria that are supported by education research theory, and therefore can encourage both highly experienced and new educators to make connections between classroom events and theoretical ideas [Amrein-Beardsley and Osborn Popp, 2011, MacIsaac et al., 2001, Morrell and Schepige, 2012]. For example, in MacIsaac et al. [2001]'s study, pre-service and inservice teachers used the Reformed-Teaching Observation Protocol (RTOP) [Sawada et al., 2002] to analyze teaching segments, and became better able to articulate what about their teaching should shift and why. The authors argue that much of this improvement can be attributed to the teachers becoming familiar with the RTOP items and better able to relate them to how science students were engaged during real events. Similarly, Morrell and Schepige [2012] explicitly take up Aubusson et al. [2010]'s framework and show evidence that their rubric-based observation protocol, the Oregon Teacher Observation Protocol (OTOP) [Wainwright et al., 2003], enabled pre-service teachers to identify conceptual and contextual anchors and led to focused, productive reflection on instruction.

As mentioned at the start of Section 4.3, parallels exist between workshop and classroom contexts that lead us to consider how workshop leaders make decisions as educators. Just as classroom educators must balance many constraints and define goals as they decide how to guide students to engage with disciplinary practices and ideas, workshop leaders must navigate their constraints and choose among potential goals as they design workshop sessions for faculty. Thus, the same attributes that characterize productive reflection on classroom instruction should inform how we support workshop leaders' reflection on workshop sessions. When it comes to improving workshop leaders' abilities to reflect, we think that a descriptive (non-rubric-based) observation tool has the greatest potential to promote fruitful, reflective discussions that are anchored to both education research theory and workshop session events without creating tension between workshop leaders and observers.

With descriptive/non-evaluative classroom observation tools, an observer captures the minute-by-minute flow of events through a set of codes that describe teacher and/or student behaviors, but does not assign a specific score to the class based on the prevalence of these codes [Smith et al., 2013, Hora et al., 2013, West et al., 2013]. This is the approach we have chosen for the R-PDOT. As the creators of the tool, we highlight aspects of workshop sessions that we consider to be strongly connected to key workshop outcomes, and in doing so implicitly create connections to education theory and our vision for professional development [Hora and Ferrare, 2013, Goodwin, 1994]. In particular, the R-PDOT data helps to portray workshop events in the way that we see them, similar to the processes of *highlighting* and *coding* that Goodwin [1994] defines as essential, often subconscious mechanisms by which people develop shared vision. In an educational context, van Es and Sherin [2002, 2008] tie Goodwin [1994]'s general depiction of developing professional vision to the ways that teachers learn to notice significant teaching events and connect them to theory, and we imagine the R-PDOT data functioning similarly here. The R-PDOT codes also employ language that is likely to bring up foundational aspects of teacher PD that could facilitate workshop leaders gaining new insights into their PD practice, as has been argued of providing rich language for teachers to reflect on their instruction [Scott et al., 2006, MacIsaac et al., 2001, MacIsaac, 2002, Wainwright et al., 2004].

Despite this implicit guidance, the descriptive nature of the tool gives workshop leaders agency to interpret the data in multiple ways and determine their own next steps, which has been shown to be important for catalyzing productive discussions about instruction [Horn and Little, 2010]. The nature of the R-PDOT data allows simple visual representations of results, which can further enable workshop leaders to make sense of the data for themselves, as we discuss in Section 4.5.2. We note that we are also well-justified in expecting the R-PDOT to serve as a research tool in future studies based on the long history of classroom observation tool use in the context of teacher education research and assessment [Hora and Ferrare, 2013, Kane and Staiger, 2011, Turpen and Finkelstein, 2009, 2010, Adamson et al., 2003, Amrein-Beardsley and Osborn Popp, 2011, Budd et al., 2013, Smith et al., 2014, Morrell et al., 2004, Walkington and Marder, 2013, Stang and Roll, 2014, Lund et al., 2015, Wainwright et al., 2004].

Many of the aspects of workshop sessions we choose to foreground with the codes themselves link back to this same literature about reflective teaching: the codes allow a user to differentiate times when faculty could be engaged in activities that mirror the productive discussions described above from times when faculty are not deeply immersed in pedagogical discussions. Specifically, R-PDOT codes can capture when faculty are analyzing concrete examples of instruction, when connections between instruction and education research might be articulated, and when multiple faculty perspectives might be voiced, debated, and discussed. We define and justify these codes in the following section.

4.4 Tool overview

Here, we define the descriptive codes that comprise the R-PDOT and elaborate on why each code was created. The R-PDOT codes broadly capture the form and focus of faculty's engagement during PD workshops, and encompass most common PD approaches (as discussed in more detail in Section 4.5). Generally speaking, the R-PDOT has a similar form and function to classroom observation tools such as the Classroom Observation Protocol for Undergraduate STEM (COPUS) [Smith et al., 2013], the Real-time Instructor Observation Tool (RIOT) [West et al., 2013], and the Teaching Dimensions Observation Protocol (TDOP) [Hora et al., 2013]: it allows an observer to collect non-evaluative, quantitative data about what faculty participants experience during workshops. An observer can simultaneously capture two complementary aspects of workshop sessions—the ways in which faculty participants are engaged and the focus of their engagement—with two sets of codes, and thus highlight aspects of workshop sessions that prior research suggests may lead to particular kinds of outcomes. Each set of codes is intended to fully span the ways that faculty are most likely to be engaged or have their attention focused during workshop sessions.

For practical reasons, the R-PDOT codes are fairly broad in scope, in the sense that a single code could be enacted in a variety of ways. By defining the R-PDOT codes, we necessarily foreground large-scale characteristics of workshop sessions that an observer can quickly record with reasonable fidelity (as justified in Section 4.5). In doing so, we lay the groundwork for reflection and follow-up analysis that targets more nuanced variations. In general, the prevalence or absence of a particular session focus, and the type of faculty engagement paired with that focus, should inform workshop leaders about what faculty outcomes are more or less plausible from a given session. For example, it would be implausible to expect a workshop session in which a workshop leader primarily lectures about education research results to improve faculty's abilities to write conceptual questions for their students; whereas a session in which faculty primarily practice writing such questions might. However, in this second example, the nature of the question-writing task, the quality of any written scaffolding the workshop leader provides, and the exact facilitation moves the workshop leader employs in the moment will also influence workshop outcomes, and these are aspects of PD environments that the R-PDOT is not designed to characterize.

As we describe the codes in this section, we elaborate on a few significant potential variations within individual codes in order to encourage readers to think critically about how each one is enacted. We also note that although we typically consider each code separately in this section, we also encourage workshop leaders to consider how codes are combined when assessing real workshop session design, as we illustrate in Section 4.6.

4.4.1 Type-of-engagement codes

The first dimension the R-PDOT allows an observer to capture is the type of faculty's engagement during a workshop session, i.e., whether faculty are listening to lecture, working independently, engaged in small group work, or engaged in some sort of large group discussion at any given time. By drawing attention to the type of engagement within faculty PD workshops and thus enabling discussions around active engagement in workshops, we clearly align our tool with the accumulated knowledge and interests of the discipline-based education research community. These codes are defined in Table 4.1, and an extended codebook (with examples) can be found in Appendix A.

Table 4.1: Type-of-engagement code names and brief descriptions.

Code name	*Code description*
Workshop Leader Lecture	Workshop leader lectures while faculty participants listen.
Large Group Closed, Faculty Participant Question	Large group closed discussion. Faculty participant(s) question the workshop leader, and (optionally) the workshop leader answers through lecturing, while all other participants listen.
Large Group Closed, Workshop Leader Question	Large group closed discussion. Workshop leader asks non-rhetorical closed questions, and (optionally) one or more faculty participants respond directly to the workshop leader [Dancy and Henderson, 2007].
Large Group Open Discuss	Large group open discussion. Workshop leader and faculty participants take turns speaking, with the discourse focused on the ideas of faculty participants [Dancy and Henderson, 2007].
Small Group Discuss	Faculty participants discuss with each other in small groups.

Table 4.1: Type-of-engagement code names and brief descriptions (continued from previous page).

Code name	*Code description*
Faculty Participant Present	One or more faculty participants present to all others.
Faculty Participant Independent Work	Faculty participants work independently on a task.

Many classroom observation tools capture aspects of participant structures that we also target [West et al., 2013, Marshall et al., 2010, Lane and Harris, 2015, Gutiérrez et al., 1999, Hora et al., 2013]. One common delineation in the participant structure-oriented literature is the distinction between *closed* and *open discussion* [O'Connor and Michaels, 1993, van Zee and Minstrell, 1997, Scott et al., 2006, Turpen and Finkelstein, 2009], which has been used explicitly by other observation tool developers [West et al., 2013, Gutiérrez et al., 1999]. *Closed discussion* refers to instances where a teacher guides students towards a single, predetermined correct answer, frequently following a pattern of I-R-E: teacher initiation, student response, and teacher evaluation [Cazden, 2001, Mehan, 1979, Lemke, 1989]. While closed discussion has the advantage of ensuring that canonical knowledge is voiced within the classroom, it can be problematic when used extensively because it positions the teacher as the sole authority and suppresses students' agency to explore, develop, and defend their own ideas [van Zee and Minstrell, 1997, Lemke, 1989, Dancy and Henderson, 2007]. In contrast, *open discussion* refers to instances where a teacher encourages student contributions while withholding their own evaluation or judgment and often asks questions that do not have predetermined answers, which supports students' development in ways that closed discussion does not.

Table 4.2: This table locates the R-PDOT type-of-engagement codes within the four classes of communicative approaches—non-interactive/authoritative, non-interactive/dialogic, interactive/authoritative, and interactive/dialogic—defined in Scott et al. [2006], Table 3, p.611. These distinctions are discussed further in the main text.

	Authoritative: purpose is to focus participants on one meaning	*Dialogic:* open to many points of view
Non-interactive: excludes contributions of other people	Workshop Leader Lecture	Faculty Participant Independent Work
Interactive: allows contributions of more than one person	Large Group Closed, Workshop Leader Question Large Group Closed, Faculty Participant Question	Faculty Participant Present Small Group Discuss Large Group Open Discuss

Researchers who have studied classroom discourse have argued that a well-balanced and well-ordered combination of different participant structures can improve learners' abilities to engage in disciplinary conversations and other practices outside of formal settings [Scott et al., 2006, Engle and Conant, 2010, Schwartz and Bransford, 1998]. We share a commitment to the importance of such outcomes when considering the engagement of any learners, undergraduate students and faculty alike, which further justifies our attention to these constructs in creating our type-of-engagement codes. When it comes to teacher PD specifically, prior research suggests that it can be productive when more experienced or pedagogically knowledgeable teachers position others as having agency in changing their own instruction instead of prescribing exact solutions [Horn and Little, 2010], which underscores the potential value of open discussion in workshop contexts.

Scott et al. [2006]'s analytical framework for distinguishing between participant structures closely matches our approach and allows us to articulate key differences between our type-of-engagement codes, as shown in Table 4.2. The authors define four classes of communicative approach using two dimensions: *authoritative* versus *dialogic*, and *non-interactive* versus *interactive*. *Authoritative* communicative approaches aim to focus participants on one meaning while *dialogic* communicative approaches are open to many points of view; *non-interactive* communicative approaches exclude contributions of other people while *interactive* communicative approaches allow the contributions of multiple people.

Within our codes, non-interactive/authoritative is exemplified by “Workshop Leader Lecture”: a workshop leader presents a single point of view and faculty participants are excluded from contributing. We also agree with Scott et al. [2006] that lecture can be dialogic in principle, meaning that a lecturer could present, compare, and contrast participants’ ideas without evaluating them, but we have found this mode of lecture to be rare within workshop settings and therefore assume lecture is authoritative unless an observer notes otherwise. In our table, we associate “Faculty Participant Independent Work” with the non-interactive/dialogic communicative approach instead. Here, each faculty participant has freedom to generate their own ideas and there is typically no opportunity for others to contribute to the development of those ideas.

Within the interactive/authoritative communicative approach, “Large Group Closed, Workshop Leader Question” maps clearly onto Scott et al. [2006]’s definition, which encompasses both I-R-E-type questioning and longer chains of questions where it is evident that the workshop leader is looking to develop one particular idea and not others. By our definition, “Large Group Closed, Faculty Participant Question” differs from “Large Group Closed, Workshop Leader Question” in that a workshop leader is answering faculty participant questions instead of the reverse. Because the workshop leader is positioned in an authority role, faculty participant questions typically lead a workshop leader to respond based on their own thinking, which maintains this participant structure as we have defined it and establishes the full exchange as authoritative. We note that these periods of addressing faculty participant questions can be minimally interactive.

The dialogic/interactive communicative approach encompasses three type-of-engagement codes: “Faculty Participant Present,” “Small Group Discuss” and “Large Group Open Discuss”. We again find Scott et al. [2006]’s approach useful to explain what we see as meaningful differences between these codes and justify these delineations. Using Scott et al. [2006]’s framework, “Faculty Participant Present” represents a *low level of interanimation of ideas*: multiple perspectives are simply made available in the public space when faculty participants present to the whole group, and there is no opportunity for consensus to be reached or differences understood if faculty voice conflicting ideas. While it could be sufficient for faculty to share ideas without discussion or debate, learning opportunities might also be missed if faculty presentation is not followed up by a communicative approach that encourages comparison. In large group open discussion and small group discussion, many ideas may be compared, contrasted, contested, and developed in relation to each other—a *high level of interanimation of ideas*. Following the same reasoning highlighted by Scott et al. [2006]’s framework, “Small Group Discuss” could have both drawbacks and affordances relative to “Large Group Open Discuss”: in “Small Group Discuss,” only a limited number of faculty participants have access to any particular group's ideas, but there are likely increased opportunities for each participant to contribute.

We return to how our real-time application of these codes compares to Scott et al. [2006]'s intent and methodology in Section 4.5.

4.4.2 Focus-of-engagement codes

4.4.2.1 Defining focus-of-engagement codes

Here we define and justify the second dimension of the R-PDOT codes: the focus of faculty participants' engagement. By focus-of-engagement, we mean the topical focus of the workshop session that faculty are asked to think about or engage with, e.g., education research results, abstract descriptions of instructional strategies, concrete examples of research-based instructional strategy implementation, or the past experiences of people within the workshop. Table 4.3 contains a complete list of the focus-of-engagement codes and brief descriptions of what they encompass. We note that unlike the type-of-engagement codes, where we expect to be able to capture the behavior of a majority of faculty participants with a single code, more than one focus-of-engagement code can occur simultaneously for faculty—these categories are not exclusive. (An extended codebook with examples of each code can be found in Appendix B, and we discuss our coding methodology further in Section 4.5.)

Table 4.3: Focus-of-engagement code names and brief descriptions.

Code name	*Code description*
Workshop Instructions	Workshop leader instructs participants about what they should or will be doing during the workshop (or participants attempt to clarify these instructions).
Education Research Theory and Results	Workshop leader and/or participants emphasize discipline-based education research processes, principles, or findings.
Instructional Strategies (IS) Description and Purpose	Workshop leader and/or participants show or describe active learning strategies ranging from current faculty practices to strongly research-based instructional strategies.
Workshop Leader (WL) Simulating IS	Faculty participants experience a workshop leader's implementation of an instructional strategy, either by acting as mock students, or through observing video, transcript, or case study narrative.
Faculty Participant (FP) Simulating IS (as educator)	A predetermined subset of faculty participants (one or more) try out implementing an instructional strategy while other participants act as mock students.
Analyzing Simulated IS	Workshop leader and/or participants reflect on (analyze, critique, evaluate, justify) a shared experience of someone simulating an instructional strategy in situ.
WL Pre-Workshop Experiences	Workshop leader and/or participants discuss a workshop leader's past experiences, including instructional goals, practices, values, and local contexts.
FP Pre-Workshop Experiences	Workshop leader and/or participants reflect on participants' past experiences, including instructional goals, practices, values, and local contexts.
Student Experiences	Workshop leader and/or participants consider students' knowledge, skills, or affect.
Disciplinary Content Knowledge	Workshop leader and/or participants consider disciplinary ideas.
Analyzing and/or Creating Student Tasks	Participants create, modify, or evaluate and/or workshop leaders critique or evaluate specific materials, questions, or tasks for students.
Planning for FP Future Teaching	Workshop leader advises and/or faculty participants plan next steps for when participants go back to their home institutions.

4.4.2.2 Justifying focus-of-engagement codes

In the remainder of this section, we consider the potential outcomes associated with enacting each focus-of-engagement code. For simplicity, we do not consider in depth what type of engagement and activity duration would plausibly generate these outcomes. However, we again emphasize that the type of faculty engagement is likely to be highly consequential to what they learn, and throughout these codes, we expect the most ambitious outcomes to emerge if there are extended periods that exemplify a dialogic/interactive communicative approach, where faculty's ideas are explored and developed. Therefore we encourage PD leaders to consider the intersection of the two R-PDOT dimensions when using this tool to inform future workshop design, as opposed to considering the potential value of the workshop foci in isolation from the extent to which faculty are engaged in making sense of relevant ideas.

Workshop instructions: A workshop leader's instructions have the potential to shape faculty participants' engagement and learning throughout the session. While faculty likely bring their own incoming expectations when they enter a workshop or session, a workshop leader can intentionally try to cue up certain ideas that might otherwise be dormant by telling faculty how to approach a particular workshop task or what they are expected to gain from attending a particular session. Hammer et al. [2005] describe the general process by which people relate their current situation to familiar past experiences as *framing* or resource activation, and several studies have shown that students' engagement and apparent abilities are influenced by the framing they adopt [Smith III et al., 1994, Disessa and Sherin, 1998, Scherr and Hammer, 2009]. Similarly, PD researchers have shown that instructors approach teaching in context-dependent ways [Markauskaite and Goodyear, 2014, Hora, 2012, Cohen et al., 2003, Stroupe, 2013, Harlow et al., 2013], which lends weight to the idea that the framing that faculty take up within workshop settings plays a role in what they gain from their participation. Although faculty may quietly or vocally contest a workshop leader's instructions, every participant will adopt some framing that influences their thinking during the session, regardless of whether or not the workshop leader guides them towards a particular orientation. Because faculty participants and workshop leaders might not naturally take up the same framing, a session in which the workshop leader's desired framing is made explicit is likely to foster stronger alignment between the workshop leader's expectations and participants' actual experiences than a session where no expectations are articulated.

Education Research Theory and Results: Unpacking the theoretical motivations that informed the development of research-based instructional strategies (RBIS) could help faculty to decide how and when to adapt, modify, or reinvent them. Prior research has shown that physics faculty often struggle to make these decisions in ways that both support positive student outcomes and fit within their local constraints, and would likely benefit from a deeper understanding of the guiding principles that underlie the developers' prescriptions [Dancy and Henderson, 2009]. More generally, faculty's ability to relate education research theory to examples of classroom practice is critical to their ability to continually assess and improve their own teaching, as discussed previously [Horn, 2010, Aubusson et al., 2010, Morrell and Schepige, 2012]. A focus on education research theory in close proximity to the "Analyzing Simulated IS" code would likely indicate that workshop leaders or faculty participants are identifying the conceptual anchors of Aubusson et al. [2010] that tie theory to practice, which we consider to be a particularly promising and rich PD activity.

Working towards an alternative goal, introducing faculty participants to education research results that are new to them might improve their ability to try innovative strategies in environments where institutional or departmental pressures constrain their teaching. In particular, physics faculty could justify their decision to use active learning strategies to resistant administrators using education research findings and methods, e.g., using concept inventories to measure student outcomes [Henderson et al., 2014, Dancy et al., 2010, Turpen et al., 2016]. That said, we caution that contrary to popular belief, quantitative research results often do not contribute to convincing individual faculty to initially try RBIS [Henderson and Dancy, 2008, Dancy et al., 2010], and we argue that a workshop that focuses on this exclusively is unlikely to shift faculty participants' willingness to try new teaching strategies.

Instructional Strategies Description and Purpose: The simplest way for faculty to become aware of RBIS is to hear them described. Prior research on the Physics and Astronomy New Faculty Workshop suggests that workshop presentations are an effective method of introducing faculty to a variety of RBIS, thus expanding faculty's awareness of what they could do in their classrooms [Henderson, 2008, Henderson et al., 2012]. Workshops could be designed to increase participants' familiarity with popular RBIS that they seem likely to take up, particularly when time is limited. Alternatively, workshop leaders could elicit instructional strategies from participants, who may contribute a greater diversity of strategies than a workshop leader would have presented based on thoughtful planning, but lead to a higher likelihood of participants attempting RBIS in their classrooms because the strategies were endorsed by their peers [Dancy et al., 2010]. Although describing instructional strategies is likely to be insufficient for preparing faculty participants to navigate the challenges of implementing them in their own classrooms, and indeed, prior research indicates many will not persist in using RBIS in the long-term after only this intervention [Henderson et al., 2012], it can give faculty a valuable starting point for experimentation within their classrooms.

Workshop Leader Simulating Instructional Strategy: Unlike describing instructional strategies in the abstract, simulating instructional strategies can give faculty participants concrete models of instruction to reflect back on. This could serve to make the nuances of RBIS implementation visible to faculty participants and more richly illustrate the possible impacts for students than a simple description [Prather and Brissenden, 2008, Turpen and Finkelstein, 2009], which may in turn increase the likelihood that some faculty will implement similar strategies when they return to their home institutions. K-12 PD designers have argued that simulating instructional strategies for participants can make PD seem more authentic and therefore make new strategies seem more plausible [Loucks-Horsley et al., 2009], and indeed, empirical evidence supports the existence of a causal link between K-12 teachers experiencing instructional strategies in workshops and later shifting their own instruction accordingly [Garet et al., 2001, Desimone et al., 2002, Darling-Hammond et al., 2009]. We note that although the three forms of simulating instructional strategies that we group together with this code—faculty participants acting as pseudo-students while workshop leaders model instructional practices; workshop leaders using video-recorded classroom episodes to engage faculty participants in classroom implementation; and faculty participants reading case studies of classroom interactions—can all provide the contextual anchors necessary for productive reflection [Aubusson et al., 2010, Horn, 2010, van Es and Sherin, 2002, 2008], each may have affordances that the others do not. We encourage future researchers and PD designers to explore the consequences of these variations through follow-up analysis, but choose not to distinguish between them in this initial classification.

Faculty Participant Simulating Instructional Strategy (as educator): Asking faculty participants to enact instructional strategies themselves could benefit them in many similar ways to what is described in the previous code, but there are also sufficient differences in the potential workshop outcomes to justify making this a separate code. In particular, while workshop leaders' simulations can provide a limited number of "expert-like" models of instruction, having many faculty participants simulate teaching strategies could provide a greater diversity of concrete, shared examples of practice that can support rich reflective discussion and debate (again, together with subsequent "Analyzing Simulated IS") [Prather and Brissenden, 2008]. Faculty may notice new aspects of these teaching strategies when simulating them as educators, and thus become better able to introspectively evaluate the fit of a particular strategy implementation to their current pedagogical beliefs and abilities. and to their local contexts.

Experiencing success at teaching during workshops could also increase faculty's self-efficacy by targeting three of the four sources of self-efficacy identified by Bandura [1986], and thus make them more willing to experiment with new strategies in their own classrooms. If faculty perceive themselves to be successful at implementing RBIS by engaging other participants in desirable ways, they could experience a sense of mastery; if faculty perceive their peers' to be successful, they could experience vicarious success; and if faculty receive encouragement or praise for their efforts as implementors, they could be persuaded to think more highly of themselves directly (social persuasion). Faculty simulating instructional strategies themselves seems more likely to improve their confidence than watching workshop leaders do so [Prather and Brissenden, 2008]: mastery experience is the most influential factor that can increase self-efficacy and cannot be achieved by watching others perform, and faculty may be more likely to experience vicarious success when watching other participants because they identify with their peers more closely.

Analyzing Simulated Instructional Strategy: As we discussed earlier, faculty's ability to analyze concrete examples of practice is a critical piece that can support continuing improvements to their instruction [Aubusson et al., 2010, Horn, 2010, Morrell and Schepige, 2012, van Es and Sherin, 2008, Sherin and van Es, 2008]. Faculty often give pedagogically superficial feedback to their peers after teaching observations, which might indicate that they also struggle to reflect on their own instruction in substantive ways. When workshop leaders vocalize their own reflective practice during workshops, they can identify alternative aspects of instruction and student engagement that may be fruitful for faculty to notice in their own classrooms. Similarly, guiding faculty to reflect on examples of instruction with their peers could improve their ability to notice consequential aspects of their students' engagement [van Es and Sherin, 2002, 2008, Goodwin, 1994, Prather and Brissenden, 2008, Darling-Hammond et al., 2009] and ultimately lead to shifts in their instruction (particularly if faculty continue to participate in similar activities following the workshop) [van Es and Sherin, 2010, Sherin and van Es, 2008]. As discussed in Section 4.3, scaffolded discussions could support faculty's engagement in conversational routines demonstrated to be productive within PD literature, e.g., focusing on specific challenges observed and connecting examples of classroom practice to general principles of teaching and learning [Horn, 2010, Aubusson et al., 2010, van Es and Sherin, 2002, 2008, 2010]. Improving faculty participants' abilities to self-assess and engage in fruitful discussions about teaching episodes could have benefits that long outlast a single PD experience, and a workshop in which this code occurs frequently could generate this outcome.

Workshop Leader Pre-Workshop Experiences: An emphasis on a workshop leader's past experiences could lead to positive and negative workshop outcomes, depending on the context and the way that individual faculty participants perceive the workshop leader. When a workshop leader effectively emphasizes their personal development as instructors, they could give faculty participants opportunities to identify ways in which the workshop leader is or was similar to them. If faculty participants see workshop leaders as role models, they might envision themselves becoming successful at implementing research-based, student-centered instruction, thus leading to an increase in their self-efficacy [Bandura, 1986], as discussed in the "Faculty Participant Simulating Instructional Strategy" code. If faculty see the workshop leaders as peers, they may become more compelled to try the workshop leader's approach to teaching than if the same RBIS were presented without an associated personal narrative [Dancy et al., 2010, Lund and Stains, 2015]. At the same time, in physics, there is evidence that that faculty who want to improve their teaching sometimes perceive education researchers to be too dogmatic and prescriptive about how RBIS "should be" implemented, which contributes to negative perceptions of the field of education research and resistance to implementing or adapting RBIS [Henderson and Dancy, 2008]. These negative perceptions could be generated or reinforced by an over-emphasis on a workshop leader's past experiences, particularly if the workshop leader portrays their instructional approach as inflexible or "correct." More broadly, if the balance between a focus on the workshop leader's and faculty participants' past experiences is strongly skewed towards the workshop leader, this could indicate that the workshop leader's ideas and experiences are being privileged over participants' ideas and experiences within the session. Workshop designers could use R-PDOT data to identify places where they could help faculty to participate more centrally in the workshop by eliciting faculty's experiences instead of exclusively sharing their own.

Faculty Participant Pre-Workshop Experiences: A focus on faculty participant's past experiences could be an effective mechanism to encourage faculty to try RBIS, and could contribute to faculty becoming more reflective teachers. For the first point, faculty are most often convinced to try RBIS by their peers [Lund and Stains, 2015, Dancy et al., 2010] and feel pressure to conform to perceived teaching norms within their local contexts [Hora, 2012, Dancy et al., 2010, Turpen et al., 2016], and faculty who are already trying out progressive instructional methods might be sought out for advice about teaching by their peers [Judson and Lawson, 2007]. Discussion surrounding faculty's own teaching could foster a sense of community among participants, which is a critical part of sustaining educational reforms [Darling-Hammond et al., 2009, Chappell, 2007, Rundquist et al., 2015], and workshop leaders could encourage faculty participants who have already tried one or more RBIS in their classrooms to share their experiences as a way to normalize the potential challenges and advantages of research-based teaching among participants [Horn, 2010]. For the second point, faculty naturally draw on a variety of prior experiences when making teaching decisions [Oleson and Hora, 2013], and it could be useful to scaffold those seeds for productive reflection by drawing them out within the workshop. For instance, if participants are encouraged to relate their past experiences to simulated classroom experiences or task design within the workshop, this could contribute to making these simulated workshop experiences seem concrete and realistic. If workshop leaders facilitate these conversations using dialogic participant structures where faculty participants voice their own ideas without the workshop leader's direct evaluation or judgment, i.e., if workshop leaders give faculty agency to recall and interpret their own experiences, these discussions could model productive conversational norms that faculty could engage in elsewhere [Horn and Little, 2010].

Student Experiences: A focus on student experiences could lead to shifts in how faculty participants perceive their own students, which again could be a positive or negative workshop outcome depending on the nature of these shifts. Faculty's perceptions of student attitudes towards active learning and academic preparation often influence their decisions to use RBIS [Lund and Stains, 2015, Turpen et al., 2016]; for instance, perceived resistance to active learning is a common barrier to physics faculty implementing Peer Instruction, as is the perception that students will not have sufficient knowledge and skills to benefit from talking with their peers [Turpen et al., 2016]. Because of these pervasive challenges, conversation surrounding student experiences could counteract or reinforce this deficit model, and potentially shift faculty's perceptions accordingly. Part of the purpose of including this code is to alert workshop leaders to when these conversations are occurring so that they can reflect on the prevailing attitudes towards students voiced within the workshop and work to direct these conversations towards understanding student perspectives. Productive conversations could help faculty think about how to teach a diverse student population in more equitable and inclusive ways, potentially through identifying how students' experiences may be racialized, gendered, or otherwise shaped by societal influences and considering how to constructively address these challenges [Seymour, 2000, Halley et al., 2011, Carlone, 2004, Steele, 1997, Johnson, 2007]. If student experiences are considered simultaneously to workshop activities coded as "Analyzing Simulated Instructional Strategies," "Analyzing and/or Creating Student Tasks," and/or "Disciplinary Content Knowledge," faculty participants may improve their abilities to identify, notice, and respond to students' disciplinary ideas [Darling-Hammond et al., 2009], which could in turn help them to assign appropriate tasks and act as effective facilitators in the classroom [Robertson, 2015, Coffey et al., 2011].

Disciplinary Content Knowledge: Workshops that incorporate content knowledge from participants' primary disciplines may be more salient to faculty than workshops that only discuss instructional strategies and learning theory generically [Loucks-Horsley et al., 2009]. Although we see disciplinary content knowledge as involving both core ideas and cross-cutting practices [National Research Council, 2012], practices often span many disciplines, and we choose to only select this code when illustrated with topics that are familiar to participants (i.e., disciplinary core ideas grounded in the disciplinary domain). Prior studies have shown that faculty's primary discipline influences how they teach and think about teaching [Lund and Stains, 2015, Singer, 1996], which suggests that a focus on disciplinary content knowledge could help faculty to perceive the teaching strategies presented at workshops as directly relevant and more easily applicable to their own teaching. When the workshop content aligns with faculty's instructional goals, they may choose to take student tasks from the workshop and use them as building blocks in their instruction. Faculty may be better able to generate new, high-quality classroom activities when the tasks they interact with during workshops target similar knowledge or skills to the content they will teach. Also, as mentioned in the previous code, faculty's abilities to identify what knowledge and skills students possess currently and could gain through instruction can importantly shape faculty's instruction [Coffey et al., 2011, Darling-Hammond et al., 2009, Robertson, 2015], and activities surrounding this (e.g., examining student work or classroom video) are most likely to be valuable when they are focused on the ideas of learners within the participants' own disciplines [van Es and Sherin, 2002].

Analyzing and/or Creating Student Tasks: Analyzing and creating student tasks within workshops could improve faculty participants' abilities to engage their students with pedagogically valuable and appropriate materials in the future. Faculty constantly go through a process of selecting, modifying, and/or creating tasks when they are planning to teach, and use a variety of criteria to determine what tasks are best [Hora, 2012, Markauskaite and Goodyear, 2014, Stroupe, 2013, Cohen et al., 2003]. The instructional materials faculty choose influence their instruction more broadly, which underscores the importance of this decision-making process [Cohen et al., 2003, Hora, 2012]. Structured workshop activities that focus on student task creation and analysis could expand faculty's vision of what criteria to consider when deciding whether and in what capacity to build from existing tasks (research-based or otherwise), and what features of classroom tasks might constrain or support student learning and engagement. Student tasks could also function as the contextual anchors that ground conversations about teaching and learning in concrete, relevant examples and thus make those conversations more focused and productive [Aubusson et al., 2010, Horn, 2010, Morrell and Schepige, 2012]. If workshop leaders scaffold faculty's productive engagement in creating and evaluating tasks within a workshop, faculty who participate in local or virtual learning communities may become better able to initiate or contribute to productive conversations within those groups later on, in addition to whatever individual gains in ability they might have made from the workshop experience alone.

Planning for Faculty Participant Future Teaching: While workshop activities represented by the other codes could help faculty participants to envision high-quality instruction and feel a sense of community within the workshop, any changes faculty are considering are more likely to be enacted if they are given opportunities to plan out next steps and think about how these changes might play out in their local contexts. Improving instruction is necessarily a long-term process, and short-term interventions like workshops can only contribute a limited amount. If faculty come into workshops with ideas about teaching and learning that are highly different from those endorsed at the workshop, they will likely struggle to achieve more desirable student outcomes without ongoing support from other educators or PD leaders. Additionally, faculty's instructional decision-making is influenced not just by how they think their students learn, but also by a host of other factors like departmental policies and incentives (or disincentives) for trying new teaching practices [Hora, 2012, Turpen et al., 2016]. Faculty may be more prepared to face these challenges if workshop leaders both encourage faculty to consider how potential changes could fit within their constraints and to identify supports (faculty learning communities, supportive colleagues, instructional materials, etc.) that could contribute to their future learning.

4.5 Methods

We now turn to the practical aspects of the R-PDOT and discuss its development, validation, and use. This section is primarily aimed at other observation tool developers and future R-PDOT users; readers who are more interested in the overall outcomes of our study may wish to skip to Section 4.6. The concept of a workshop observation tool emerged from private communications with Dr. Edward Prather, and as did our connections to the research settings that supported the tool's development. We attended many iterations of two of the largest disciplinary workshops for faculty, the Physics and Astronomy New Faculty Workshop [Henderson, 2008] (6 iterations) and the Center for Astronomy Education Tier I Teaching Excellence Workshop [Prather and Brissenden, 2008] (3 iterations), as we developed the R-PDOT codes. We used published descriptions of K-12 and faculty PD [Garet et al., 2001, Loucks-Horsley et al., 2009, Darling-Hammond et al., 2009, Council of Scientific Society Presidents, 2012] to ensure that the codes span the activities most likely to occur in any workshop aimed at helping faculty to improve their teaching.

The functional form of the R-PDOT is an online interface, hosted on the Tools For Evidence-Based Action community's General Observation and Reflection Platform (GORP), which allows an observer to code by selecting and deselecting buttons continuously throughout a "live" workshop session[1]. Consistent with the nature of the codes, an observer can select one type-of-engagement code and one or more focus-of-engagement code(s) at any given time. Observers can also record timestamped comments that complement or explain these coding decisions. We note that in order to select a particular code, an observer may need to assume that there is more shared purpose in the room than may actually be the case. For instance, during many types of engagement (most notably lecture and closed discussions), the only publicly visible evidence of faculty participants' thinking is discourse between

[1]The R-PDOT interface can be accessed at http://gorp.ucdavis.edu.

a limited number of speaking actors. In the case of no counter-evidence, we default to maintaining the workshop leader's framing. We base counter-evidence for the focus-of-engagement on a small subset of participants in the observer's vicinity, and shift to other codes or add codes as appropriate. Along similar lines, faculty's type-of-engagement may conflict with the workshop leader's instructions, e.g., faculty may spontaneously begin to discuss a prompt with their peers when asked to work independently, in which case we scan the room and select the code that matches the most common participant behavior. Otherwise, we maintain the type-of-engagement established by the workshop leader.

We find several of Scott et al. [2006]'s guidelines for labeling communicative approaches useful in further helping observers to identify code shifts. In particular, we agree that taking an overview [Bakhtin, 1986] is an appropriate orientation to determining what is happening in a workshop session, as all turns of talk are linked to other turns of talk and not isolated from what was occurring previously. Taking an overview can include both considering the overall teaching purpose when deciding whether or not a code is occurring, and using contextual, non-verbal cues to help establish whether or not an interaction is evaluative. This approach is particularly relevant in determining the duration of a particular focus-of-engagement and deciding the type-of-engagement during large group interactions, where potential ambiguities can be resolved by considering how the session has been unfolding. We can often make inferences about the type-of-engagement in a large group interaction based on an initial question statement, e.g., a workshop leader may pose an obviously open or closed question to the group, and, as above, we assume this type-of-engagement is maintained unless there is evidence otherwise.

By default, our coding methods ultimately diverge from Scott et al. [2006]'s recommendations because the R-PDOT is intended to enable users to capture workshop events in real-time (either through in situ observations or watching video played without pause), not to provide a framework for slow, iterative coding of video episodes. We agree with Scott et al. [2006] that it is easier to identify the nature of various interactions through replaying video multiple times, and sometimes do this ourselves when qualitatively analyzing a workshop session in depth. Some of the boundaries we have created are easier to capture in a first pass than the boundaries in Scott et al. [2006]'s framework; for example, "Workshop Leader Lecture" is a single code, even though, as we noted earlier, lecture could be authoritative or dialogic in principle. More fundamentally, however, the value of enabling users to capture these aspects of workshop sessions "live" and without multiple iterations of coding so that workshop leaders and evaluators can immediately reflect on workshop sessions, greatly outweighs the benefits of doing so with the greatest possible accuracy. Although we have strived to make the distinctions between codes as transparent as possible to R-PDOT users, we acknowledge this underlying limitation as we explore the resulting inter-rater reliability below.

4.5.1 Establishing inter-rater reliability

Establishing inter-rater reliability (IRR) requires both selecting metrics to assess reliability relative to common standards set by other observation tools, and finalizing a set of codes for which to measure reliability between observers. In the following section, we describe how two IRR metrics—the Jaccard similarity score and Cohen's κ—are appropriate for and can be applied to the R-PDOT coding scheme; the ways in which we ensured that the current code names and descriptions are accessible to potential users; and the results of IRR testing between the first and second authors. We also consider the circumstances under which future users may (or may not) wish to establish strong IRR with the R-PDOT for themselves.

4.5.1.1 The Jaccard similarity score

The first metric we use to calculate IRR, the Jaccard similarity score, is simply the fractional observed agreement between two observers who have coded the same set(s) of data. For example, if observer A and observer B code the same workshop session and agree 85% of the time about whether or not lecture is occurring, the Jaccard similarity score for "WL Lecture" would be 0.85. We calculate the Jaccard similarity score for each code individually to allow comparisons to the reliability of the Classroom Observation Protocol for Undergraduate STEM (COPUS) instrument [Smith et al., 2013], which has a similar functional form and has undergone extensive reliability testing. The standard formula for the Jaccard similarity score, applied in this context, is provided in Smith et al. [2013], and we follow their strategy for IRR testing wherever possible throughout this section.

One notable difference between Smith et al. [2013]'s approach and ours is the exact method of data collection. With the R-PDOT, observers collected data continuously throughout an observation—keeping relevant codes selected until a change in the session occurred, while Smith et al. [2013] collected data in 2-minute intervals—checking off all codes that occurred in a given time interval. Our method of data collection provides a more accurate visual and quantitative description of workshop sessions; however, in order to calculate IRR, it is necessary to bin the data. We initially chose to bin into 1-second intervals for simplicity and to ensure sufficient sampling fidelity. In doing so, we recognize that it is unrealistic to expect observers to simultaneously change codes at that level of accuracy, i.e., within the same second, particularly when they may be selecting two or more appropriate codes based on the same change in the workshop session. We therefore decided to incorporate some additional leniency into our IRR calculation, so that minor differences in the timing of selecting codes do not reduce the overall IRR scores.

In order to determine what amount of leniency is appropriate for this coding method, we used an excerpt of workshop video to quantify how many seconds of delay between observers can typically be attributed to realistic lags in observers noticing and selecting new codes instead of real disagreement about which codes describe a particular workshop event. To calibrate this, the first author coded the same 7-minute video segment two times consecutively before any other IRR scores had been calculated, and the two sets of data were compared using the Jaccard similarity score as described above (as if the data were collected by two observers). The video segment was selected to be straightforward to code but containing many transitions between codes, such that any differences between the two iterations could be attributed to natural delays in selecting codes rather than the observer's uncertainty about code definitions. Consistent with this assumption, the two iterations of coding were quite similar initially: all Jaccard similarity scores were 0.97 or higher. For each code, we then omit any data within x seconds of when that code was selected or deselected from the inter-rater reliability calculation, increasing x until all codes had similarity scores of 1.0, i.e., perfect agreement. In this way, we found that omitting data that fell 2 seconds before and 2 seconds after the selection or deselection of a particular code was sufficient to bring the two observations into perfect agreement. We therefore exclude data that is 2 seconds or less away from an observed code shift in all subsequent IRR calculations.

4.5.1.2 Cohen's kappa

While the Jaccard similarity score has the benefit of being easier to interpret than Cohen's κ and always possible to calculate, it is not predictive of how well two observers will agree in the future and thus is not a true measurement of IRR. Cohen's κ, on the other hand, takes into account the fact that two observers might agree by random chance in any given observation, and therefore yields a conservative, statistically driven estimate of how likely it is that they will agree in the future.

The formula for Cohen's κ is:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where κ is the probability that the two observers will agree in the future; p_o is the observed proportionate agreement between the two observers; and p_e is the probability of random agreement between the two observers, based on how often codes were selected by each observer. (The expressions for p_e are described below.)

We calculate Cohen's κ for each focus-of-engagement code separately. This is justified because observers can select several focus-of-engagement codes at once, which implies observers must decide whether a particular code should be selected largely independently of whether or not other codes are selected. This calculation of Cohen's κ draws on our previous calculation directly—p_o is equivalent to the Jaccard similarity score—and mirrors the calculations described in Smith et al. [2013]. In particular, like Smith et al. [2013], we consider agreement for a given focus-of-engagement code to indicate times when both observers have selected the same code, and times when both observers have chosen not to select that code. Here, p_e for a particular code is given by the expression:

$$
\begin{aligned}
p_e = & \frac{\text{\# of times code was selected by observer A}}{\text{\# of times A selected any focus-of-engagement code}} \\
& \times \frac{\text{\# of times code was selected by observer B}}{\text{\# of times B selected any focus-of-engagement code}} \\
& + \frac{\text{\# of times code was not selected by observer A}}{\text{\# of times A selected any focus-of-engagement code}} \\
& \times \frac{\text{\# of times code was not selected by observer B}}{\text{\# of times B selected any focus-of-engagement code}}
\end{aligned}
$$

where only instances where both observers are actively coding (i.e., have selected some combination of type-of-engagement and focus-of-engagement codes) are included in these calculations.

The type-of-engagement coding differs from the focus-of-engagement coding in that observers are choosing between a set of mutually exclusive codes, i.e., there will always be exactly one type-of-engagement code selected at a given time. Because of this, we calculate a single κ for all of the type-of-engagement codes together. In this case, it is unnecessary to include times when two observers agree that a particular code is *not* selected when calculating Cohen's κ; in other words, only times when both observers agree that a particular type-of-engagement code is selected count in this calculation. Thus the expressions for p_o and p_e are slightly different. Here, p_o is the total number of instances when the two observers select the same type-of-engagement, and p_e, the probability of random agreement, is given by the expression:

$$p_e = \sum_i \frac{\#\text{ of times code } i \text{ was selected by observer A}}{\#\text{ of times A selected any type-of-engagement code}}$$

$$\times \frac{\#\text{ of times code } i \text{ was selected by observer B}}{\#\text{ of times B selected any type-of-engagement code}}$$

We note that because of its probabilistic nature, Cohen's κ is only useful when there is some variation in the coding of at least one observer. If there is no variation in the coding considered for a given calculation, then κ is undefined. Similarly, if there is only variation in one observer's coding, e.g., if one observer selects a single code for an entire session, or if one observer never selects a particular focus-of-engagement code, then κ is automatically zero for the relevant calculation. This is a logical result that indicates a lack of statistical information needed to make a reliable predictive model: all of the observed agreement is taken to be random agreement. More generally, when there is little variation in observers' coding, either because a code was selected highly frequently or highly infrequently by one or both observers, the formal probability that observers will agree by random chance is quite high. Therefore we expect κ values to be low in those cases even when Jaccard similarity scores indicate a close agreement.

4.5.1.3 Refining codes

As we were defining the R-PDOT codes, we periodically sought and incorporated feedback from potential users in order to ensure that the version of the R-PDOT coding scheme, for which we calculate IRR, is accessible and comprehensible to users. Dr. Stephanie Chasteen, the NSF external evaluator for this project and an experienced PD presenter, pilot-tested our preliminary focus-of-engagement codes "live" during a workshop by informally noting examples of each code and identifying points of confusion. We made several significant changes as a result of this feedback, including adding new codes, rewriting code names, and modifying code meanings. A tenured Astronomy faculty member, PD participant, and recent adopter of RBIS, Dr. Derek Richardson, also provided feedback on comprehension and code names when the codes were closer to their final form. In response, we modified language and code names to make them more accessible to users with a wide variety of expertise in education research.

To further improve and validate the accessibility of the R-PDOT, we conducted one-on-one interviews with workshop leaders surrounding R-PDOT data. We interviewed six presenters from the Physics and Astronomy New Faculty Workshop (NFW), incrementally introducing them to code descriptions, anonymized example data from NFW sessions, and data from their own sessions when possible. We prompted them to talk through their interpretations, and offered our own ideas when necessary to create a productive and collaborative interview setting. These interviews confirmed the general accessibility of the code names and descriptions: workshop leaders were typically able to interpret the codes meanings with little or no guidance from the interviewer. We made minor changes to some code descriptions based on one interview, which reduced the number of words in the description without changing the meanings of the codes. More substantive points of confusion raised during these interviews could be better addressed through modifications to the visual output of the tool, as discussed further below. In addition to contributing to the development and validation of the R-PDOT, these interviews provide preliminary evidence that the R-PDOT can indeed support workshop leader reflection as intended, and we describe these results briefly in Section 4.7.

4.5.1.4 IRR results

Over the course of this study, we video-recorded 64, 45-minute to 1-hour long sessions across three iterations of the NFW. Once we had established a final set of code names and descriptions, we (re)coded multiple NFW video segments to demonstrate formal IRR for the R-PDOT coding scheme. (We had coded and discussed both "live" and video data periodically throughout this project in order to refine this set of codes and establish joint understanding of code meanings.)

The first and second author independently coded 7, 5–30 minute segments of video based on the final R-PDOT codebook, which totaled 98 minutes of data (Table 4.4). The first author used her preliminary coding to select video that spans a range of workshop activities and participant structures. For each segment, she indicated a key phrase that indicated the start of the segment, the initial codes to select (since selecting a button on the R-PDOT interface marks the start of data collection), and the duration of the episode for coding. To limit biases in coding, the first and second authors coded all of these video segments independently before comparing to each other, and only the first author reviewed the preliminary coding data to choose the video segments.

When selecting video data for this exercise, the goal was to include video segments such that each focus-of-engagement and type-of-engagement code would be present in the final, combined IRR dataset, and such that each focus-of-engagement and type-of-engagement would be similarly prevalent to allow for more direct comparison of reliability measures across codes. Despite these ideals, however, practical considerations of how the type and focus of faculty's engagement played out during the NFW created natural limitations to how evenly we were able to sample different codes. In particular, because "WL Lecture" was highly prevalent across the NFW, it was difficult to avoid oversampling this type-of-engagement relative to other types while still capturing a range of focus-of-engagement codes, as evident in Table 4.4. Similarly, "Workshop Instructions" and "Planning for FP Future Teaching" were quite rare, and we were the least able to sample these out of the focus-of-engagement codes. (The limitation in our ability to capture extended enactment of the “Workshop Instructions” code will likely persist in future studies due to the nature of that code. We discuss the lack of planning for future teaching and the pervasiveness of lecture in our analysis.)

We find that our Jaccard similarity scores for every code are all 85%, or above which we consider to be sufficient for this study. We note that the standard set by the COPUS team [Smith et al., 2013] is for raters to achieve Jaccard similarity scores above 90%. The κ values we include represent a lower limit on the likelihood that these two observers will agree in the future, and provides another metric for future R-PDOT observers to compare to. Landis and Koch [1977] provide some general guidance for how to interpret these results: they characterize κ values as poor (<0.00), slight (0.00–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–1.00). By these benchmarks, we find that IRR for the type-of-engagement codes is moderate (0.56). We also find that the two observers simultaneously selected focus-of-engagement codes occurring for more than 10% of the time with fair to moderate IRR, and that only focus-of-engagement codes selected for less than 5% of the total time showed poor or slight IRR. Because those codes occurred so rarely, we should not be surprised that the reliability in selecting them was worse. Thus, we consider these results to be sensible and sufficient for our purposes, but note that these rare codes may need more refinement in the future if/when it is more possible to capture them.

Because we do not have immediate plans to explore quantitative research questions with the R-PDOT ourselves, we choose not to iterate among ourselves for higher reliability nor to pursue IRR with additional observers at this stage. As the initial developers of the R-PDOT, we primarily aim to outline the procedures needed to perform IRR calculations for this instrument in order to streamline this for future researchers. While we have established what we consider to be a viable and useful set of descriptive codes and have demonstrated that it can be used reliably within our research team, we recognize that future researchers may wish to modify the codes as the R-PDOT is tested in other settings, which would require IRR to be revisited. To the extent that future users wish to use the R-PDOT as a reflective tool only, establishing IRR and maintaining these exact codes may not be necessary: in these cases, we consider it more important that the observer and observee agree about code meanings. The Generalized Observation and Reflection Platform (GORP) where the R-PDOT is currently hosted allows users to easily create custom versions of the R-PDOT for their own use and share these custom versions with others. Based our experiences interacting with users of classroom observation tools like the COPUS, it is reasonable to anticipate that some custom modifications and improvements could be made over time.

Table 4.4: Results of inter-rater reliability testing for the authors, measured with the Jaccard similarity score in all cases and Cohen's κ when appropriate. These calculations include 98 minutes of data comprised of 7, 5–30 minute segments of NFW sessions. Results for codes that were selected for more than 10% of the total time coded are in bold, and results for codes that were selected for less than 5% of the time are italicized. As justified in Section 4.5.1, we calculate a single value of Cohen's κ the type-of-engagement codes, and separate values for each focus-of-engagement code.

Code name	*Minutes coded (observer average)*	*Fraction of total time (observer average)*	*Jaccard score*	*Cohen's κ*
WL Lecture	**61.0**	**0.62**	**0.85**	–
LG Closed FPQ	**13.1**	**0.13**	**0.93**	—
LG Closed WLQ	7.4	0.08	0.91	—
LG Open Discuss	*2.8*	*0.03*	*0.96*	—
SG Discuss	6.1	0.06	0.98	—
FP Present	*4.9*	*0.05*	*0.97*	—
FP Ind Work	*2.6*	*0.03*	*0.98*	—
All type-of-engagement codes	**98.0**	**1.0**	—	**0.56**
Workshop Instructions	*2.8*	*0.03*	*0.96*	*0.07*
Education Research Theory and Results	**21.1**	**0.22**	**0.93**	**0.77**
IS Description and Purpose	**31.9**	**0.33**	**0.85**	**0.64**
WL Simulating IS	**11.4**	**0.12**	**0.90**	**0.41**
FP Simulating IS	8.6	0.09	0.97	0.75

Continued on next page...

Code name	*Minutes coded (observer average)*	*Fraction of total time (observer average)*	*Jaccard score*	*Cohen's* κ
Analyzing Simulated IS	8.4	0.09	0.89	0.22
WL Pre-Workshop Experiences	**11.9**	**0.12**	**0.88**	**0.36**
FP Pre-Workshop Experiences	5.9	0.06	0.95	0.44
Student Experiences	5.9	0.09	0.89	0.18
Disciplinary Content Knowledge	**23.6**	**0.24**	**0.87**	**0.60**
Analyzing and Creating Student Tasks	*4.1*	*0.04*	*0.97*	*0.58*
Planning for FP Future Teaching	*1.7*	*0.02*	*0.97*	*-0.01*

4.5.2 Data visualization

We consider it a critical piece of the R-PDOT development to link the numerical output described above to accessible, intuitive visuals. We intend the R-PDOT output to serve as a reflective anchor for workshop leaders, as well as a mechanism for researchers to identify compelling video data for future analysis quickly and easily, and this necessitates visualization design choices that take into account the nature of desirable sense-making around R-PDOT data. In this section, we describe the representations of R-PDOT data we have developed, tested, and refined based on user feedback.

The most basic representations of the R-PDOT output are summative plots showing the percentage of the total session time spent enacting each code. We take a similar approach to the designers of the COPUS tool in this respect [Smith et al., 2013]; however, we agree with Lund et al. [2015] that although pie charts are easy to interpret quickly, they are misleading when the data represented can sum to more than 100%. Because our type-of-engagement codes represent mutually exclusive categories and our focus-of-engagement codes do not, we use pie charts to represent the prevalence of the type-of-engagement codes and bar charts to represent the prevalence of the focus-of-engagement codes.

While we initially repeated the same color palette across the two sets of codes, in the final version, we have associated each code with a unique color, and shown in the color key (Figure 4.1). Although this change increased the total number of colors in our representations, we found that workshop leaders consistently sought to asso-

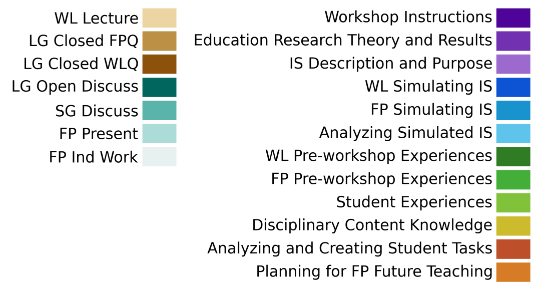

Figure 4.1: Color key for the type-of-engagement (left) and focus-of-engagement (right) R-PDOT codes.

ciate codes with like colors across categories during interviews, e.g., relating a yellow focus-of-engagement code to a yellow type-of-engagement code. This mapping was unintentional on our part, and we do not wish to narrow workshop leaders' vision for how activities could be structured by implicitly encouraging particular associations. Furthermore, we found that workshop leaders were sometimes overwhelmed by the overall number of codes, and confused by which color associations were intentional and which were arbitrary. Although we have included a greater number of colors in our final version, we now only use similar colors when highlighting similarities between codes, thus simplifying the interpretive work for the user.

The final color palette for the type-of-engagement codes, based on a ColorBrewer palette for diverging data[2], cleanly maps onto the four communicative approaches as outlined in Table 4.2. Brown in our color scheme is associated with authoritative communicative approaches while teal is associated with dialogic approaches, and darker shades of the same hue indicate an equal or greater level of

[2]The original ColorBrewer (v1.0) was funded by the NSF Digital Government program during 2001–02, and was designed at the GeoVISTA Center at Penn State (National Science Foundation Grant No. 9983451, 9983459, 9983461). The design and rebuilding of the new version (v2.0) was donated by Axis Maps LLC, winter 2009 and updated in 2013.

interactivity compared to neighboring codes. We cluster the focus-of-engagement codes using a separate palette with five hues—purple, blue, green, yellow, and orange—and up to three different shades (and therefore codes) per hue. While the associations within the focus-of-engagement clusters are not as robust as the type-of-engagement associations in the sense that similarities in the research underpinnings are weaker, these clusters can help to initially orient the user to what kinds of activities are occurring and therefore reduce cognitive load when interpreting data. Specifically,

- *purple* indicates codes that relate to describing or making sense of established knowledge;
- *blue* indicates codes that relate to in situ simulation of instructional strategies;
- *green* indicates codes that relate to the experiences of various stakeholders in the educational process;
- *yellow* indicates disciplinary content knowledge (a single code); and
- *orange* indicates codes that relate to primarily forward-looking activities.

We consider the connections among the blue and green codes to be the strongest, and note that comparing the prevalence of the codes within these clusters adds to the interpretive power of the tool. These colors match the colors on the online R-PDOT interface, which allows coders to develop facility in interpreting these plots during the data collection process.

While the summative plots provide an overview of workshop sessions, which can be useful for comparing across multiple sessions and inferring potential session goals, they do not showcase the full range of information captured with the

R-PDOT. Inspired by the design decisions made by West et al. [2013] for the Real-Time Instructor Observation Tool (RIOT), we also created sets of timelines to represent R-PDOT data.[3] These timelines enable users to develop interpretations that rely on coordinating the type-of-engagement with the focus-of-engagement or understanding how a sequence of events unfolded. Before timelines were introduced during interviews, more than one workshop leader stated that whether or not lecture was broken up by other types of engagement mattered to their assessment, and this information is clearly visible in a timeline representation. Our interviews with workshop leaders also support the idea that timelines have particular affordances for helping presenters to reflect on their own sessions. We suspect that workshop leaders are better able to recall recent events and generate specific modifications to implement in future sessions when looking at timeline visualizations instead of an overview alone. For instance, in considering data from their own session (coded by the interviewer), one workshop leader stated

> *"I like the timelines because I know what slides were happening, and so seeing how those translate into these categories I think is interesting. I could see myself, if I'm giving a lot of workshops, going through and saying, (pointing to codes on the focus-of-engagement timeline) okay this activity is hitting these bits, what happens if I do an activity that hits these bits? And then seeing what happens, what's the outcome. Because*

[3] Currently, the Python scripts used to generate these timelines and all other representations shown can be obtained on the Digital Repository at the University of Maryland (DRUM) or by contacting the lead author. Ultimately, we anticipate that these visualizations will be generated automatically following data collection on the GORP website, and the original Python scripts will be downloadable from a GORP repository.

now I know I'm not doing this. So maybe I should try it and see what happens. And then I can map it into what I was actually doing (motioning to the type-of-engagement timeline)."

As researchers who study PD workshops through multiple methods, we have also found these timelines useful in helping us to quickly select video segments for both detailed qualitative analysis (e.g., Olmstead and Turpen [2015]), as a representation akin to a "content log" in some respects, and IRR testing, as alluded to earlier.

Visually, the timelines maintain the same color scheme as the summative plots. Because the type-of-engagement codes again represent mutually exclusive categories while the focus-of-engagement codes can occur simultaneously to each other, we show the occurrence of type-of-engagement codes on a single, color-coded timeline and the focus-of-engagement codes on stacked timelines, with a single code in each row. We note that the type-of-engagement color palette we adapted from ColorBrewer is colorblind safe; thus, we can use all of these colors on a single row without limiting access to colorblind users.

We have selected three sessions from the Physics and Astronomy New Faculty Workshop (NFW) as examples of how these visualizations appear, and to demonstrate what inferences can be made from them. Out of the 64 NFW sessions we video-recorded, we chose two sessions that are fairly typical of the NFW based on our experience and preliminary coding, and re-coded them from video using the finalized R-PDOT codes. These sessions allow us to illustrate some affordances and limitations of common faculty PD approaches, and will likely seem familiar to workshop leaders. We also selected a third session that contains design elements we highly value but find to be rare within the NFW. By including this session, we hope to illustrate the capacity of the R-PDOT to document a wide range of session types and to expand what workshop designers see as possible within short faculty PD sessions. We note that while we hypothesize about possible outcomes based on the empirical and theoretical work we have already introduced, we do not claim that any of these outcomes have occurred. Instead, we simply aim to illustrate the capacity of the tool to help workshop leaders and researchers to develop plausibility arguments for session outcomes and make to judgments about what post-workshop data would help to diagnose these outcomes. Our interpretations of these three workshop session follow in the next section.

4.6 Session analysis

The R-PDOT data for three sessions of the NFW—labeled Sessions A, B, and C—are shown in Figure 4.2. Sessions similar to A and B occur regularly at the NFW, while C is less typical. Each session lasted for approximately one hour (A: 60 minutes; B: 65 minutes; C: 48 minutes). We provide our interpretations of each session based on these plots in the text that follows.

4.6.1 Session A

Type-of-engagement: Workshop leader A used authoritative communicative approaches (shown in brown) for more than 75% of the session time. Pre-planned

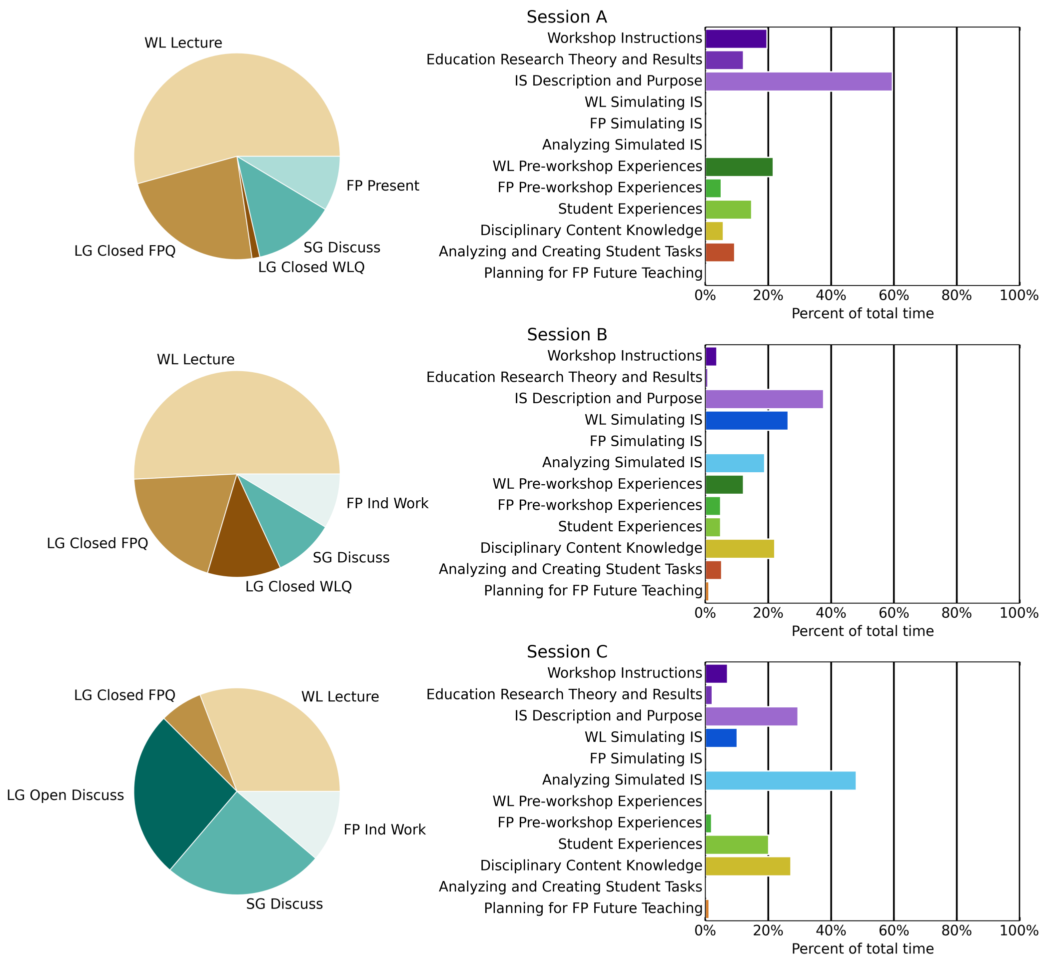

Figure 4.2: Summative plots showing how much time was spent enacting each R-PDOT code during three approximately 1-hour long sessions of the Physics and Astronomy New Faculty Workshop, labeled A, B, and C. The pie charts show the type of faculty's engagement as defined in Table 4.1, where authoritative communicative approaches are shown in brown and dialogic approaches are shown in teal; the bar charts shows the focus of faculty's engagement as defined in Table 4.3. Further discussion of the color scheme can be found in Section 4.5.2.

lecture accounted for most of this, followed by the workshop leader responding to faculty participant questions. For the $\sim 20\%$ of the session that was dialogic (shown in teal), most of it took the form of small group discussion, which allowed participants to share and develop their ideas with a few other participants, and having faculty present ideas to the whole group (potentially following this small group discussion).

Focus-of-engagement: A majority the session was spent focusing on previously established knowledge, particularly the description and purpose of instructional strategies ($\sim 60\%$ of the time). A small amount of this discussion of established knowledge focused on education research theory and results, likely to justify or support the use of these instructional strategies. A minimal percentage of the time was spent engaging in analysis of student experiences ($\sim 15\%$) and student tasks ($\sim 10\%$), less than the percentage of time spent considering workshop instructions ($\sim 20\%$) that may have framed those activities. The workshop leader also shared his/her personal experiences during the session, but did little to elicit or discuss participants' prior experiences.

Potential outcomes: Faculty would likely leave this session more aware of the RBIS discussed, and potentially motivated to use the strategies if they seemed to address pressing concerns or needs. It seems unlikely that faculty participants would gain more than a surface-level understanding of these strategies, however, and would likely not take up new ideas that conflict with their current views about teaching and learning. We claim this in part because the session was so lecture-heavy, which gave faculty highly limited opportunities to voice, reason through, or refine their

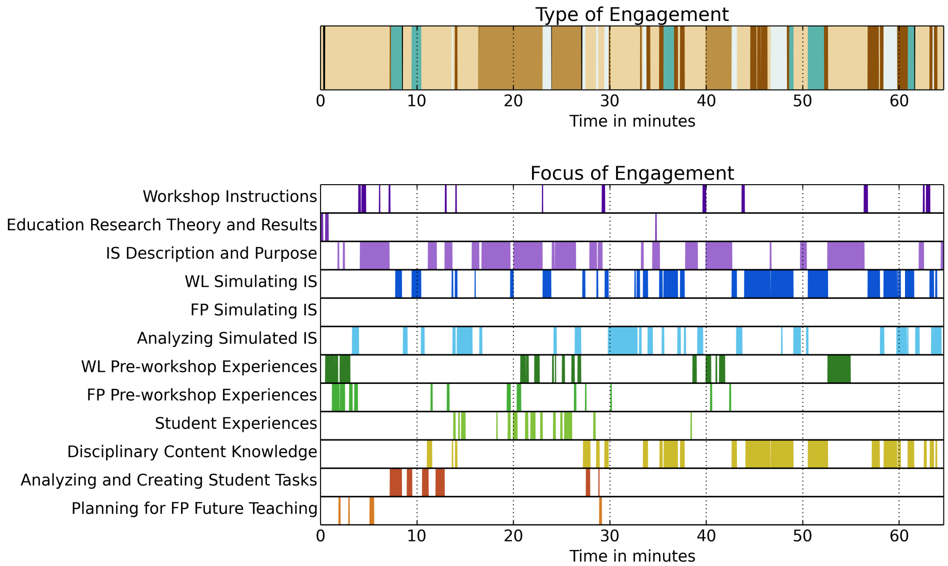

Figure 4.3: Timeline representations of the R-PDOT data for Session B, showing how the type (top) and focus (bottom) of faculty's engagement shifted during the session. The colors are consistent with those in Figures 4.1 and 4.2, where again, brown indicates authoritative communicative approaches and teal indicates dialogic communicative approaches for the type of engagement, and the focus-of-engagement colors indicate loose clusters.

current ideas. Moreover, regardless of the type of engagement, the whole session was primarily oriented towards abstract descriptions of strategies, with highly limited attempts to connect these strategies to concrete experiences within or outside of the workshop.

4.6.2 Session B

Type-of-engagement: The workshop leader in Session B used authoritative communicative approaches (brown) for over 75% of the time, as in Session A, mostly by lecturing and then by answering faculty participants' questions. Faculty were given time to think through ideas independently and could have explored divergent ideas in small groups, while all large group discussion was authoritative or closed.

Focus-of-engagement: Although Session B still primarily focuses on the description and purpose of instructional strategies, with some justification through research studies, the focus-of-engagement profile differs significantly from what we saw in Session A. Here, the workshop leader draws participants' focus to the concrete implementation of instructional strategies, and gives faculty the opportunity to experience these strategies as pseudo-students. There is less emphasis on the workshop leaders' past experiences, and some (albeit limited) mention of faculty and students' experiences.

Coordinating codes: Because the focus of faculty's engagement is highly variable (especially in comparison to Session A) and seems to target more than declarative knowledge, we can learn more by considering a timeline representation showing how the events in Session B unfolded (Figure 4.3). In carefully examining these timelines, we make two key observations about the nature of the interactions in the session: first, that all of the dialogic types of engagement occurred when faculty were experiencing well-defined RBIS, typically in conjunction with disciplinary ideas, and second, that all of the analysis of the in situ teaching events was dictated by the workshop leader, either through lecture or short periods of closed questioning. We also find that although the description and purpose of instructional strategies is the most prevalent code, the workshop leader simulates instructional strategies and analyzes their implementation intermittently throughout the entire session, indicating to us that this was a consistent theme and likely salient to participants.

Potential outcomes: Modeling instruction for faculty participants may help them to envision changes to their practice that would be difficult to identify in the abstract, and therefore this session could be more successful than Session A at shifting faculty's instruction towards a student-centered model. The workshop

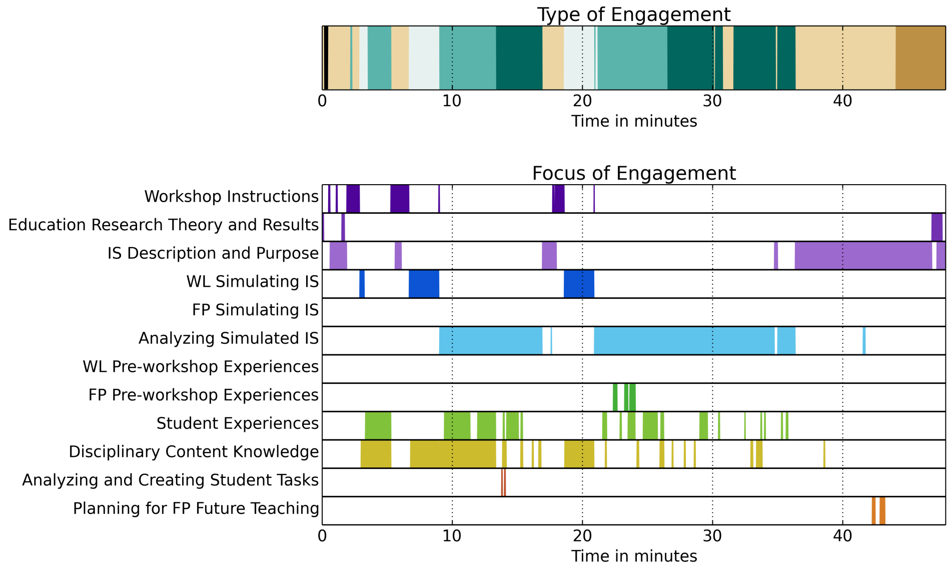

Figure 4.4: Timeline representations of the R-PDOT data for Session C.

leader demonstrated for faculty how they could justify the instructional choices they experienced as students, which might improve faculty's ability to reason about these specific strategies. We anticipate that faculty would likely make well-reasoned choices about whether to adopt or reject the workshop leaders' suggestions based on their incoming pedagogical understanding and how well their goals align with the workshop leader's justifications. Some faculty participants might also use the workshop leader's reflection on their practice as a model for how to engage in reflective practices outside the classroom more generally. But because the workshop leader presented these analyses of instruction through authoritative communicative approaches, there were limited opportunities for faculty to develop their abilities along this dimension within the session.

4.6.3 Session C

Type-of-engagement: The type of engagement in Session C is mostly dialogic ($\sim 60\%$), including a quarter of the session time spent with participants engaged in both large group open discussion and small group discussion. There are certainly some authoritative segments in Session C, primarily lecture and some responding to faculty questions, but the structure of the session overall dramatically contrasts with the authoritative approaches of workshop leaders A and B. This distribution of types of engagement seems to indicate that workshop leader C is primarily focused on helping faculty participants to develop their own ideas as opposed to prescribing a single approach.

Focus-of-engagement: The primary activity across the entire session is analyzing simulated instructional strategies. Student experiences are included in this discussion, whereas the workshop leaders' experiences are never considered and faculty's experiences are considered very rarely. As in Session B, the time that workshop leader spends focusing on the description and purpose of instructional strategies is greatly reduced relative to Session A, and instead, the workshop leader focuses faculty participants on concrete examples of instructional strategies simulated in situ.

Coordinating codes: Looking at the timelines for Session C (Figure 4.4) provides information about how these activities are structured and coordinated. We can see that the workshop leaders' simulation of instructional strategies happens in two discrete chunks with faculty participants working individually, which suggests that faculty may be watching a scenario being enacted as opposed to participating in it as pseudo-students. In contrast to Session B, faculty are made responsible for the complex task of analyzing the simulated instructional strategies: all of the reflection and analysis occurs within interactive, dialogic participant structures. We also note that the final 10–15 minutes of Session C take on a different structure and focus than the rest of the session, namely, the workshop leader describing instructional strategies through lecture and answering faculty questions.

Potential outcomes: The outcomes we might expect from Session C are quite different from Sessions A and B. Because we see faculty first being given agency to analyze instruction with a small subset their peers, and then further developing and comparing the many ideas that may have emerged in these small group discussions with the workshop leader's subtle guidance, we find it highly plausible that this session could improve faculty's ability to reason about new pedagogical situations, notice new aspects of their students' thinking, and consider a diversity of potential teaching choices. Faculty might also become better able to initiate and engage in these kinds of pedagogical discussions with other educators in the future. On the other hand, Session C does not link directly to faculty's prior teaching experiences or project into their future teaching. These factors combined with the lack of a prescribed instructional model (as in Session B) may make it difficult for faculty to identify concrete next steps for improving their instruction based on this session. The dialogic nature of this session may also challenge faculty's expectations about the workshop session, e.g., faculty who expect that they will be told how to teach may perceive the session to be less valuable, and result in lower evaluations compared to more prescriptive sessions.[4]

4.7 Discussion

The information captured with the R-PDOT can be used to consider the merits and limitations of sessions taken individually, as demonstrated in the previous section, and to consider workshop design more holistically. For the two sessions we present as representative of many others within the Physics and Astronomy New Faculty Workshop (NFW) (Sessions A and B), we find that dialogic types of engagement are highly limited, and that presenters rely heavily on lecture. In contrast, dialogic types of engagement are highly prevalent in the one session (Session C) that is anomalous within the NFW. These three sessions also illustrate that although opportunities for faculty consider the description and purpose of instructional strategies occur regularly in NFW sessions, opportunities for faculty to interactively experience concrete examples of practice are a central focus in some sessions (like Sessions B and C) but highly limited or non-existent in others (like Session A). We consider lecturing about instructional strategies likely to be a successful mechanism for raising faculty's awareness of what RBIS exist, which is highly consistent with prior research on NFW outcomes [Henderson et al., 2012], and we consider these increases in awareness to be beneficial to an extent. In particular, we think that faculty who already possess extensive pedagogical knowledge or participate in faculty learning communities could most easily take advantage of new resources in ways that gen-

[4] On a post-workshop evaluation survey, faculty's ratings of the "usefulness" of Session C was lower than for many other sessions at that iteration of the NFW.

uinely enhance their students' learning and engagement. On the other hand, we find it difficult to imagine how faculty who are less knowledgeable or experienced with research-based teaching would productively assess and enact new instructional strategies based on sessions that do not allow them to interact with these strategies in some way. Because of this, we consider concrete, interactive examples to be a critical component of PD workshops that can enable all faculty to clearly envision changes to their instruction, and have shown how we can identify these instances with the R-PDOT.

The R-PDOT data can also reveal important aspects of whether and how these concrete examples of practice, when present, are explored and assessed within workshop sessions. Both Sessions B and C spend significant time analyzing instructional strategies simulated in situ, which likely indicates that these workshop leaders have goals beyond just raising faculty's awareness of RBIS. Of the two sessions, only Session B models reflective teaching, while only Session C extensively engages faculty in evaluating teaching decisions for themselves. We find significant value in workshop leaders modeling productive reflection because they are likely adept at weaving education research theory into their analysis and noticing key events. But we argue that it is also important for faculty to have some experiences generating and refining pedagogical ideas for themselves (i.e., as learners about teaching) if we expect them to enact these dialogic/interactive approaches within their own classrooms. Moreover, as stated earlier, we do not consider the benefits of a balance between authoritative and dialogic communicative approaches to apply exclusively to teaching science content, and we consider guiding faculty to make well-reasoned but autonomous decisions an important characteristic of workshops that could contribute to faculty's long-term success. Our data suggests that workshop leader C's approach may be rare within faculty PD workshops, and we argue that this potential tendency to strongly skew towards authoritative communicative approaches warrants consideration from workshop leaders in the future.

The lack of occurrence of focus-of-engagement codes can also inform workshop design. While it would likely be unproductive to attempt to enact all of the focus-of-engagement codes in a single 1-hour session, individual PD facilitators could generate potential modifications to their own sessions based on codes that never occur, and PD organizers could consider what new sessions would best complement existing ones within a longer workshop. For example, when looking across these three sessions, we see that the "Planning for Faculty Participant Future Teaching" and "Faculty Participants Simulating Instructional Strategies" codes never occur, and the "Education Research Theory and Results," "Faculty Pre-Workshop Experiences," and "Analyzing and Creating Student Tasks" codes occur rarely. These gaps primarily relate to activities that could help faculty to fit RBIS into their local contexts, which is known to be a limitation of faculty PD more broadly [Dancy and Henderson, 2009]. Encouragingly, sessions that target these activities have been added to the NFW while we have been conducting this research, and other disciplinary faculty workshops enact them as well [Council of Scientific Society Presidents, 2012]. But it is unclear to us how central these activities are or will become for participants in each of these workshops, and how different disciplinary workshops compare to each other. With the R-PDOT, PD leaders can gather concrete evidence to use as they discuss ways to continually improve their workshops, such as ways to incorporate more design structures that could prepare faculty to adapt instructional strategies for their unique situations and motivate faculty to participate in long-term PD efforts.

Through our interviews with workshop leaders, we have found preliminary evidence that R-PDOT data can indeed support others' engagement in the type of reflection and analysis that we have demonstrated here. For example, we find that looking at R-PDOT data encourages workshop leaders to think about and articulate potential limitations of lecture. For example, while reflecting on some example sessions from the NFW (similar to Sessions A and B), one workshop leader states:

> *"I think the limitation [of lecture]—I think there's shared experiences, there's experiences that each person has in the classroom or in the group that others can learn from. And here we're supposing that the person at the front is the expert and knows that, and they very well may be, but there are some experiences that can probably add greatly to the learning of everybody who's participating. And so if there's not a lot of facilitation of that engagement of the broader group then you could miss out on those experiences. And things sometimes organically grow, you know, sometimes the conversation goes in a place that you never thought it would go and it's an even better experience for everybody."*

We have also found that workshop leaders may generate and justify potential changes to their own sessions based on codes they are not yet enacting. In reflecting on their own session after reading through the R-PDOT codes and seeing example data, one workshop leader realized how to incorporate pedagogical ideas from their physics instruction into their workshop session, stating:

> *"One of the things that I would do differently that it would take the exact same amount of time is what I actually do in my classes, which is, for [a specific physics prompt], instead of asking them individually [i.e., only calling on individual participants], to have them think about it for a minute and then talk amongst themselves, because instead of hearing from one person that there were two ways of doing it, likely working together they would be less afraid—just like our students—of responding, and likely come up with two or three [ways] and be less afraid to respond. ??? I can tell you from looking at the taxonomy just as with when I've looked at [the] RTOP and the COPUS, that you sit there and you go 'Oh if I—' and it's like 'Oh it's just a minor change to make it more interactive or different.' "*

4.8 Conclusions

As a community of researchers who want to help bring about improvements in undergraduate STEM education, we need to critically evaluate our own professional development practices for faculty. Our observational tool, the Real-time Professional Development Observation Tool (R-PDOT), contributes to that goal by illuminating and allowing others to document key PD practices currently used within workshops aimed at improving faculty's instruction. The R-PDOT can give a holistic sense of how interactive and prescriptive workshops are by capturing the ways in which faculty are engaged and what they are focused on. Initial empirical results from the Physics and Astronomy New Faculty Workshop suggest that faculty PD can be lecture-heavy, non-interactive, and authoritative, leaving little space for faculty to grapple with their incoming ideas about teaching and learning or their instructional contexts. If we want workshops to contribute to the pursuit of ambitious learning goals for faculty, beyond building initial awareness and interest, further shifts in PD practices will be needed.

We intend our discussion of PD activities in this paper to serve a broad audience. In exploring these PD activities and ways of engaging faculty in thinking about teaching and learning, we hope to expand our community's vision for what faculty PD could look like. In addition to drawing attention to the interactivity and prescriptiveness of workshop sessions, we encourage increased attention to forward-looking activities (i.e., analysis of student tasks, analysis of pedagogical approaches, and planning for future instruction), lending weight to promising features of PD that are already being enacted in some PD settings. For workshop leaders who choose to take up the tool to document their own practices, the R-PDOT output can generate critical, reflective discussions surrounding aspects of PD that are likely to be consequential for workshop outcomes, and thus meaningfully inform workshop leaders' future efforts. The R-PDOT can also fuel future research and inspire us to ask new questions about how we can best support faculty learning, such as "What does it look like for faculty to productively engage in developing course content together?," "How do faculty PD workshops vary across disciplines, or when workshop facilitators have different kinds of expertise?," and "How do facilitators elicit and help faculty to build from their own experiences?" For researchers, the R-PDOT can provide baseline documentation that would enable the full pursuit of such questions.

Methodologically, the R-PDOT could be used to support future research in a variety of ways. In quantitative research, the R-PDOT could allow correlations to be made between workshop outcomes captured by survey, interview, or classroom observation data and workshop design when used across multiple sessions or workshops, thus furthering our understanding of how these design decisions are consequential to faculty. For example, the R-PDOT data highlights unanswered questions about the importance of incorporating disciplinary content knowledge into PD, which could be explored by searching for correlations between the prevalence of this code and relevant survey responses. In qualitative research, the R-PDOT data can lead users to ask targeted questions about the details of workshop design and implementation that are related to but not captured by these broad codes, such as questions about workshop leader facilitation moves or faculty interactions. These questions could be pursued by using the R-PDOT to identify episodes for detailed analysis, and coordinating with additional data sources such as workshop video, field notes, or workshop artifacts.

Taking a broader perspective, researchers could also continue to pursue how and to what extent R-PDOT data acts as formative feedback for workshop leaders through additional interviews, discussions, and continued documentation. Such research could strengthen efforts to use observation tools as a catalyst for teacher reflection within PD and elsewhere, and help to unpack some current orientations to faculty PD in ways that could inform ongoing conversations about how members of the discipline-based education research community can effectively contribute to increasing the impact of research-based teaching innovations.

Chapter 5: Case studies: linking workshop design to faculty's engagement

5.1 Introduction

Throughout this thesis, we have argued that part of what faculty need to develop in order to be successful in their classrooms is the ability to reason about instruction in ways that better align with education research principles, and to make well-informed decisions about if, when, and how to adopt or adapt research-based instructional strategies (RBIS) within their local contexts. This could be described as developing "professional vision" [Goodwin, 1994], or learning to see the classroom in ways that are more consistent with the collective knowledge and insights of the education research community. We have also argued that we should treat faculty as partners in instructional change, and that improving faculty's abilities to reason about pedagogy would enable them to add to existing instructional change initiatives in more useful ways. We thus set out to examine PD workshops with the intent of drawing attention to whether and how we might be promoting these kinds of ambitious faculty learning outcomes within workshop sessions.

The question of how we are supporting faculty's learning in workshops is inherently one of design. In Chapter 2, we suggested that Sandoval [2014]'s analytical framework of "conjecture mapping" for designing learning environments in general provides a useful approach to conceptualizing and researching workshop design. Sandoval [2014] states that design can be separated into embodiment (structures that shape the learning environment), mediating processes (learners' interactions in these environments), and outcomes (how learners change as a result of their experiences). We agree with Sandoval [2014] that proposing and iteratively refining hypotheses about how embodiment is linked to mediating processes, and how mediating processes are linked to outcomes, presents a direct, well-defined path towards understanding and ultimately improving the design of learning environments.

In both Chapters 2 and 4, we considered other researchers' recommendations about key structural features that can support teacher learning (as well as learning more generally) as a starting point for thinking about PD workshop design. In particular, we suggested that creating high similarity between faculty's local contexts and workshop contexts, e.g., by bringing authentic problems of practice into workshop settings, could create opportunities for faculty to explore potential everyday routines in an environment that provides support and guidance and thus might increase the likelihood of faculty improving their instruction. In Chapter 4, we drew attention to several structural features, i.e., elements of workshop embodiment, related to these activities by naming them as codes in our workshop observation tool, the R-PDOT. As we defined these codes and showed example R-PDOT documentation of workshop sessions, we proposed a wide variety of learning opportunities that might arise as a result of these varied design structures. While developing the

R-PDOT at the Physics and Astronomy New Faculty Workshop (NFW), we noticed that opportunities where faculty's ideas about instruction might be debated and discussed were quite rare. We were able to document this lack of opportunity easily with the R-PDOT, which underscores its utility for initiating evidence-based discussions about workshop design.

While R-PDOT data provides a useful starting point for studying workshop design, there are intrinsic limitations to what this relatively broad, real-time documentation can tell us, as we have acknowledged previously. For instance, R-PDOT data can indicate when open and small group discussions occur, but it does not guarantee that productive conversations are happening when these codes are selected, nor it does not show how productive conversations can be supported or launched. Similarly, we noted that the presence of the "analyzing simulated instructional strategies" code implies that pedagogically-rich ideas (e.g., justifications of instructional moves based on plausible student behaviors) are being voiced, but we were unable to state with confidence that the ideas were indeed pedagogically rich, nor to illuminate in detail how conversations coinciding with this code unfolded. We suggested that fine-grained variations in workshop implementation could be consequential for faculty's engagement in workshops and thus for workshop outcomes, and that identifying consequential variations could be better approached through more thorough, iterative analysis.

In this chapter, we will go beyond what the R-PDOT documentation can capture to pursue a deeper understanding of how faculty interact during workshop sessions, and what design structures seem to shape these interactions, by analyzing workshop video. In Sandoval [2014]'s language, we aim to illustrate mediating processes and make evidence-based hypotheses about what specific aspects of workshop embodiment may lead to these mediating processes. In developing these "design conjectures," we can imagine how other constructs we introduced in Chapter 2 might re-emerge as useful here. In particular, recalling our discussion of theoretical frameworks, we note that considering the nature of the "resources" that faculty draw on during their interactions may direct our attention to consequential design features. For example, we suggested that the framing that is cued up for faculty will likely color their interactions by changing what ideas and behaviors are readily available to them, and that a particular workshop activity, such as acting as pseudo-students or critiquing other's instruction, could be framed in multiple ways. In this analysis, our theoretical and analytical approach will allow us to consider faculty's actions and talk in light of the framing that the workshop facilitators seem to cue up, instead of assuming that faculty have only one way they could approach these situations (which would likely seem more constraining for design). Similarly, we will be able to consider the nature of the tasks, materials, instructional strategies, etc., that faculty interact with in our analysis. We suggest that these factors could both contribute to cueing up particular ideas or behaviors and serve as resources (in a broader sense of the word) that become part of faculty's thinking and learning. Flexibly considering a range of possible resources and contextual cues will allow us to develop more accurate hypotheses about what causes faculty's interactions to play out in the ways we observe and use these hypotheses to point workshop designers in potentially fruitful directions. Moreover, by capturing the nature of faculty's interactions, we will also be better able to hypothesize about what outcomes are plausible from these particular sessions than we were in Chapter 4. Thus, following this analysis, we will be better able to communicate with workshop leaders about why they may (or may not) want to interact with faculty in particular ways when enacting potentially worthwhile activities, and provide initial evidence of the challenges and affordances of various structural elements for supporting the kinds of interactions we value based on the possibility that they would lead to ambitious PD outcomes.

5.2 "Doing school" versus sense-making

Before we embark on this detailed, qualitative analysis of workshop sessions, we will first take the time to review literature that is newly relevant in this section. In the previous chapter, the literature on how students and teachers interact in the classroom enhanced our ability to describe the structure or form of faculty's participation in workshop sessions in a broad sense (e.g., by giving us language to describe and justify the "type-of-engagement" codes). Here again, considering what we know about students' interactions and behaviors in classroom settings can set us up to understand how faculty might engage in workshop sessions. In this case, some of the potential links we will highlight are more direct, in the sense that we are not only looking at general interactional patterns that could occur in any learning environment, but we are also anticipating that faculty may recall and repeat interactional norms that they experienced as students [Lortie and Clement, 1975]. In workshops, we are treating faculty as learners in a structured setting, sometimes asking them to engage in collaborative tasks, and sometimes even directly asking them to act as if they are science students. Thus, as we suggested in Chapter 2, it is not unreasonable to think that workshop settings could cue up faculty's ideas and intuitions about how to act in the classroom based on their recollections of how they experienced the classroom as students. Because most faculty likely experienced largely "traditional" instructional contexts, their past experiences of being students are likely well-captured within the literature that we will summarize.

5.2.1 "Doing school"

A broad characterization of how many students experience their academic trajectories has been labeled as "doing school" [Pope, 2001, Jimenez-Aleixandre et al., 2000] and has been problematized on multiple scales [Pope, 2001, Jimenez-Aleixandre et al., 2000, Turpen and Finkelstein, 2010, Lemke, 1989, Barron, 2000]. On a large scale, it has been demonstrated that because of the cultural and societal pressures students face to be competitive and excel in school, the "best and brightest" students learn to manipulate the school system in various ways in order to "manage their workloads" and get ahead, typically at the expense of their own physical and emotional discomfort [Pope, 2001]. They form alliances with teachers and other authorities, "cheat" in various ways, multi-task (e.g., do classwork for one class during another), and argue for higher grades. These students learn to adapt to different academic settings by figuring out what will be rewarded in each, for instance, by knowing when they will be praised for asking questions even if those questions are superficial and functionally meaningless. Even though these students often prefer to be engaged in meaningful learning, they feel these behaviors are necessary to be successful and meet other's expectations, and therefore fall into routines that we call "doing school."

While "doing school" can refer to students' orientations to their whole academic experience and is dependent on a breadth of cultural factors, this culture is partly generated and/or recreated within science classrooms, and other researchers have looked more narrowly at how "doing school" manifests itself in these spaces [Jimenez-Aleixandre et al., 2000, Turpen and Finkelstein, 2010, Lemke, 1989, Barron, 2000]. For example, Jimenez-Aleixandre et al. [2000] shows how students "doing school" or "doing the lesson" when engaging in a scientific task may involve enacting assigned procedures without thinking about the purpose of these procedures or steps relative to the substance of the task. These assigned steps may not always be the most appropriate steps for the scientific task at hand, and the scientific task itself may not be particularly substantive or meaningful. However, students recognize what procedures they are "supposed" to follow to meet the instructor's expectations, and do so. In these kinds of activities, students often accurately see completing the required steps as sufficient for receiving the instructor's approval, and thus do not pursue scientific ideas or ask questions that might naturally emerge from their experimental findings in these classroom settings.

In interactions between students and instructors, frequent use of the initiate-respond-evaluate (I-R-E) pattern of discourse that we described in Section 4.4.1 also tends to be the way that instructors enforce school-like norms [Lemke, 1989]. In particular, instructors may use I-R-E with the aim of "get[ting] students to know the scientifically accepted answers the scientists have developed to describe the natural world" but at the expense of "build[ing] sensible and plausible models of the natural world that are intelligible to the students themselves" [Turpen and Finkelstein, 2010]. Turpen and Finkelstein [2010] describe this distinction as *answer-making* versus *sense-making*, where answer-making, i.e., focusing discussion exclusively on canonically correct answers, language, and explanations, is a part of what comprises doing school. These school-like norms can also negatively influence how students interact with their peers when they work in small groups. In particular, Barron [2000] identifies specific conversational markers such as students taking independent solution paths, violating turn-taking norms (students talking over each other), and conflicts of insistence (students arguing that they are correct without inviting their logic to be interrogated). Barron [2000] suggests, and we agree, that these and other markers of "low coordination" in group work are in part driven by a culture that encourages competition and individual accomplishment over fostering shared understanding among students. Moreover, we can see how a focus on correct answers and following pre-determined procedures without sense-making would discourage students from trying to develop deeper understanding with their peers in this collaborative work. Thus, many aspects of students' classroom experiences could be negatively impacted when school-like norms are at play.

We imagine that school-like norms could emerge during workshop sessions both because of the framing that workshop leaders portray and because of faculty's incoming expectations about what and how they will learn in this setting. In particular, even though faculty often choose to adapt instructional strategies within their local contexts, they might expect to be told prescribed instructional solutions by workshop leaders who are have been positioned as experts by virtue of being invited to present (and sometimes by their involvement or leadership in the development of these instructional strategies), which could contribute to faculty acting in ways that resemble "doing school."

5.2.2 Sense-making

We can see how faculty taking up school-like framings in workshop might limit workshop outcomes and conflict with the ambitious goals we laid out. However, we do not think this is the only orientation that faculty could adopt. We maintain that faculty could (and should) be supported in what we will call pedagogical sense-making, e.g., evaluating instructional choices based on observations or plausible hypotheses about students' thinking and behavior, during workshops. We already explored some potential characteristics of pedagogical sense-making when we described productive conversational routines in Section 4.3, such as drawing on concrete examples of practice ("contextual anchors"), linking instructional choices to education research principles ("conceptual anchors"), and discussing and debating choices and hypotheses with other instructors [Aubusson et al., 2010]. Given the challenges and limitations that traditional school-like norms can create in the classroom, researchers have also articulated and explored alternative classroom norms that would be preferable for students, and we note that desirable faculty conversations would align with many of these norms. In particular, as we alluded to in the previous subsection, many K-16 researchers also argue that sense-making (in this case about science ideas) and related constructs are a critical part of this desirable alternative to doing school [Jimenez-Aleixandre et al., 2000, Engle and Conant, 2010, Turpen and Finkelstein, 2010, Coffey et al., 2011]. For example, Jimenez-Aleixandre et al. [2000] argue that argumentation is central to learning, i.e., classroom instruction should focus on encouraging students to develop and defend evidence-based and theoretically-motivated claims. They also argue that this process of argumentation should become the norm, such that students recognize that they are expected to engage in these conversations with their peers and begin to do so naturally. Similarly, Engle and Conant [2010] argue that fostering student-driven debate and reasoning about authentic scientific questions in ways that mirror professional scientific practices, what they call "productive disciplinary engagement," is more important than communicating predetermined disciplinary content to students. We see strong connections between this goal for science students and the PD goal of helping faculty to develop "professional vision," which further suggests that we are justified in drawing similarities between classroom and workshop contexts in these ways.

As in the case of "doing school," K-16 researchers have articulated some instructional and interactional markers that would allow sense-making to emerge in the classroom. Discursively, we have already broadly outlined ways that instructors could interact with students that would be consistent with sense-making when we defined "open discussion" and "high interanimation of ideas" in Section 4.4.1. Here, instead of exclusively enacting I-R-E, instructors might foster sense-making by asking open-ended questions and use responsive teaching moves (e.g., Brodie [2011], Richards [2013], Robertson [2015]) to draw out learners' reasoning and encourage them to respond to their peers' reasoning. Sense-making could also play out in students' small group interactions. In particular, we again find that Barron [2000] presents a useful set of markers of student behavior, in this case, markers of high coordination in group work. In contrast to the low coordination markers we listed in the previous subsection, markers of high coordination include co-construction of ideas (students work to build shared meaning), respecting turn-taking norms (students listen to each other), and productive conflicts (students interrogate each other's ideas when disagreements arise). Just as markers of low coordination seemed tied to school-like norms, here, we see that these markers of high coordination would allow students to construct logical claims together and engage in productive discussions.

We anticipate that faculty will likely participate in workshop sessions following either or both of these contrasting norms: they may act in ways that are consistent with how they might have behaved in school—going through the motions without a particular purpose beyond appeasing the instructor—or in ways that are consistent with how faculty might (or could begin to) approach problems they encounter in their own classrooms—critically analyzing pedagogical choices and considering alternatives. Thus, this literature provides a foundation for analyzing faculty's engagement workshop sessions, and we will consider potential distinctions between "doing school" and sense-making in the analysis that follows.

5.3 Methods

In total, we video-recorded 63, 45-60 minute sessions across 3 iterations of the NFW. Because we are particularly interested in whether and how faculty reason about RBIS, and because sessions that focused on well-established RBIS seem to have the potential to support either pedagogical sense-making or "doing school," we restricted our video selection to strategy-specific sessions (51 sessions). We also selected from video in the later two workshop iterations of our dataset (42 sessions, 36 focused on RBIS) when our data collection process had become more refined, allowing us to better capture faculty's talk. Of these 36 sessions, we again narrowed our criteria to include only sessions that contained significant faculty talk, and used preliminary R-PDOT coding (and occasionally field notes) to identify episodes of extensive small group and/or open discussion, which limited our data to 15 sessions. We re-watched and re-coded 12 of these sessions, taking notes on the nature of faculty interactions and talk, stopping when we had identified a sufficient number of episodes that spanned a particular kind of faculty experience that aligned with our research interests. Specifically, we chose episodes that involve faculty participating in the behaviors captured by the "Analyzing Simulated Instructional Strategies" R-PDOT code, since we consider this code to be a potential marker of pedagogical sense-making. We note that this selection criteria excluded sessions where R-PDOT codes such as "Creating and Analyzing Student Tasks" and "Student Experiences" better captured the dialogic interactions that occurred within the session. Based on this initial coding, we think these sessions also have the potential to support ambitious PD outcomes, and future analysis of these sessions could reveal other kinds of pedagogical sense-making.

At the NFW, analysis of simulated instruction was launched by faculty experiencing concrete examples of research-based instruction in three different contexts: faculty acting as instructors (the "Faculty Participant Simulating Instructional Strategies" code), faculty acting as pseudo-students (the "Workshop Leader Simulating Instructional Strategies" code), and faculty watching classroom video (also the "Workshop Leader Simulating Instructional Strategies" code). Where possible, we selected two cases for each context to allow us to develop design conjectures about how design structures are linked to faculty's engagement in each situation by identifying meaningful differences. As a result, we ultimately selected 5 cases from 6 sessions:

– two cases where faculty act as pseudo-instructors,
– two cases where faculty act as pseudo-students, and
– one case where faculty watch classroom video.

For this last context, only one instance of faculty watching and analyzing classroom video together occurred within our complete dataset, and therefore our ability to make evidence-based conjectures about how to support faculty learning in these kinds of sessions will be more limited.

In each of these cases, we describe the general flow of the session and analyze transcript at points when these activities are occurring, i.e., when we can observe how faculty talk and act. In cases where faculty's talk is prevalent throughout the session, we select cases that seem to accurately represent faculty's experiences in the session (as opposed to being idiosyncratic or anomalous). When we analyze faculty's small group interactions, we consider video of the one focal group we had recorded for that particular session. While these small group interactions likely do not represent the experiences of all faculty in the session, they are likely not unique either, and thus provide useful information of what kinds of interactions these session designs can support. In this particular analysis, the only small group interactions we analyze occur in the cases where faculty act as pseudo-students while the workshop leader simulates the role of the instructor.

As we analyze workshop video in this chapter, we pursue the following research questions: (1) In what ways are faculty's interactions during workshop sessions consistent with "doing school" and/or pedagogical sense-making? (2) What structural features, particularly workshop leader facilitation moves, seem to contribute to cueing up or sustaining "doing school" and/or pedagogical sense-making? In particular, as we have suggested previously, whether faculty's behaviors are consistent with these traditional school-like norms or pedagogical sense-making will likely be contingent on how workshop sessions are structured and facilitated. In our analysis of contrasting cases, we will be able to identify differences in workshop facilitation that seem to influence the framing faculty adopt. When we analyze faculty's interactions with each other, we will draw on Barron [2000]'s markers of high and low coordination in group work, as well as literature on teacher's conversational routines, to help characterize these interactions and ground our observations in prior research. While we will not centrally pursue analysis of how the particular tasks or other structural features of the NFW are consequential for faculty's engagement, we will note these possible influences as we look across cases in our final discussion.

Before launching into this analysis, we note that we use the following transcript conventions throughout this section. In sessions where we analyze large group interactions, we identify faculty by letters that are assigned alphabetically according to when the faculty participants first spoke. If it was not possible to easily identify the speaker, or when many faculty speak at once, faculty participants are identified as "FP" or "FPs." The workshop leaders are identified as "WL," with a subsequent identifier such as "-1" or "-2" when needed. When transcribing speech, we adopt the conventions given in Table 3.1 of [Ochs, 1979] to denote the start and end of overlapping speech: "\\ ]" and the start and end of tentative transcriptions where it was difficult to hear the speaker: "( )." Descriptions of gestures and other non-verbal cues are enclosed in brackets.

5.4 Faculty's pseudo-instructor experiences

5.4.1 Overview

In this section, we consider two parallel sessions in which faculty practice the implementation of a particular RBIS. We choose to begin with the analysis of these sessions in part because they present the most highly coordinated contrasting cases in our dataset: the workshop leaders had communicated with each other about the session structure ahead of time and followed the same general design template, and because these sessions occurred at the same point during the workshop, both groups of faculty participants bring similar shared experiences from prior sessions.

Specifically, these sessions center on teams of faculty trying out the implementation of Think-Pair-Share (TPS). The design of this 1-hour session and a related 1-hour session are based on the "situated apprenticeship" PD model developed by Prather et al. [2009], which has been routinely implemented and refined at the Center for Astronomy Education Tier I Teaching Excellence Workshops and regional teaching exchanges for several years. In an earlier NFW session, groups of 3–5 faculty were assigned a physics topic (e.g., rotational motion, work and kinetic energy, etc.) and asked to write a conceptually rich, multiple-choice question that targeted a particular student reasoning difficulty. They were also told that they would be implementing TPS the following day using the question they had written, and that they should prepare and coordinate for implementation with their group once their question was completed. A workshop leader had modeled this instructional strategy for them in previous sessions and elaborated on the pedagogical value of specific steps and phrases that he used. Faculty were also given a "rubric" (see Appendix C) which draws attention to the steps of TPS [Prather and Brissenden, 2008]. Some of the rubric questions are targeted at specific scripted phrases, such as "Did the presenter ask 'Do you need more time?' before going to the first vote?," while others are potentially more open-ended, such as "Did the presenter appropriately direct the students to engage in discourse about their answer choices and explain their reasoning using a prompt that would foster an active discussion?"

In this set of sessions, the question-writing groups collectively act as the instructor in implementing TPS, while other faculty act as pseudo-students. The group of faculty who are acting as the instructor divide up this role between them, for instance, one group member might direct participants to vote on the question for the first time, then pass the instructor role to another group member who would direct participants to discuss with their peers, and so on. Within the mock implementation, faculty participants are asked to imagine that a mini-lecture on the topic had just been given. As Prather et al. [2009] describe, faculty participants who are not acting as the instructor are also asked to act as critical friends for the presenters. In faculty's role as critical friends, they are encouraged to "pause" the session at any point to critique the TPS implementation. The workshop leaders also watch the presenters and "pause" the implementation themselves, but strive to fade out of this process over time by asking faculty to fill in the reason for the "pause." At the end of each TPS round, faculty are given a short amount of time to critique the question itself. We note that the two sessions we analyze may not well-represent the model described in Prather et al. [2009] because of the strong time constraints placed on this activity in the NFW as compared to the time constraints in a typical CAE workshop, and because these sessions are embedded in a workshop that may foster different faculty expectations and behaviors than the CAE workshop.

Here, we highlight three parallel episodes in each of two different implementations (labeled B and C) of this "same" workshop session: (1) the workshop leader's initial framing of the session, (2) the first team's implementation of TPS and the critique of this implementation, and (3) a conversation that emerged organically in both sessions about the potential affordances and drawbacks of using the exact same instructional script in every class.

5.4.2 Initial framing

These two episodes mark the start of each session, where each workshop leader (WL), WL-B and WL-C, describes the session structure and their expectations for participants. We begin with Session B.

5.4.2.1 Session B

WL-B: Hi, welcome to class. I need to lay some ground rules. We are going to have one group up presenting. You've presumably practiced. I'm going to believe all of you did your homework and you've decided on who is doing what part of implementation. When there is a group up there presenting, it is our job—not mine, yours—to "pause" them if they have missed a step in the implementation that (a WL) demonstrated and that you had a link provided for a "how to" guide on. If you pause someone, don't look at me and wait for me to say, "Well why did you pause them?" You paused them. You have a reason. Just say, "Pause, you should have blah blah blah." Okay? Because we don't want to waste any more time on that.

There are two reasons why critiquing is important: it's important for the people who are presenting, so that they have the time here in a loving and supportive and caring environment, in the safety of others, to practice, stumble, take the missteps and try it over again, get it right, before doing it in front of students. That is a very powerful thing for you to be able to go home with. You will feel more comfortable and secure in trying this in your class when you go back home.

For those of us that are doing the critiquing, you are at your highest level of understanding something when you are able to synthesize what's happening and to assess and evaluate it. So that's what you're supposed to be doing. If I don't hear you pausing people, I'll pause you and I might just randomly point at someone and say, "You say why I paused them." That's uncomfortable. I don't want to do that. It's uncomfortable for all of us. Usually what happens if I'm having to "pause" you is that you are spending too much time thinking about physics. That's your focus, is getting the question right. That's not the reason why we are here. The reason why we're here is to be students for our teachers and to be critical friends for our colleagues. Any questions before we get going?

WL-B explicitly frames the session as school-like, and positions faculty as students within the workshop. She includes phrases that denote school—"welcome to class," "you've presumably practiced," and "I'm going to believe you all did your homework"—which establishes some attributes that would characterize a good "student" coming into the session. It is implied that a correct implementation closely tracks what another workshop leader modeled and what is in the written guide. WL-B makes a point to separate faculty's responsibilities from hers—"it is our job (not mine, yours) to "pause" them if they have missed a step in the implementation"—which also might cue up traditional student-teacher distinctions and imply that faculty and WL-B are not partners in evaluating instructional decisions. She illustrates what it might look like for faculty to not follow the rules (not giving an explanation after saying "pause," not "pausing" frequently enough), with an implication that faculty might do this, might be reprimanded for it ("Well, why did you pause them?," "You say why I paused them"), and might feel uncomfortable as a result. WL-B also hypothesizes that some anticipated negative behaviors (lack of participation) would likely indicate that faculty are thinking about the wrong things, i.e., "you are spending too much time thinking about physics." A risk of "wasting time" also comes up here, which suggests that WL-B's perceptions of strong time constraints might weigh heavily in her facilitation choices.

WL-B also articulates what kinds of session norms she wants to establish, and states what the benefits to faculty will be. These norms are for faculty to feel able to make mistakes and correct them based on friendly but critical feedback from their peers, so that they will feel more comfortable when they implement TPS in their own classrooms. We note that in order for faculty to easily take up this framing, they would need to share WL-B's goal to "get it right" within the session (implying that there is a correct way of implementing TPS that works in all contexts), and plan to adopt this pedagogical strategy in their own classrooms.

5.4.2.2 Session C

WL-C: This is Think-Pair-Share. You should be in whatever room you were in for that yesterday. Does everybody have a rubric?
FP: How do we know whether we have a rubric?
WL-C: Rubric says, "Rubric for Think-Pair-Share implementation." I want at least one in each group. You can share.
FP: I don't think I have this. I don't have it.
WL-C: Does everybody in your group have one? If you don't have one in your group I can give you one. Everybody found their group? Awesome.
D: Are we filling these out or something?
WL-C: No. Here's what's going to happen. We have 10 minutes for each group and we're already 9 minutes behind. I will keep us tightly to the clock. What will happen is that for example, simple harmonic motion will come up here and they will present their question and you guys need to use the rubric to help you remember when to yell "pause." They're going to do their Think-Pair-Share question and you should yell "pause" whenever these questions, whenever we're not all doing the optimal job that we can do. There will be 8 minutes for the presentation, there will be 8 minutes on implementation and then in the last two minutes we'll talk about the content of the question. That may feel a little weird because you spent so much time thinking about that content last time. But this is all about the implementation. I will yell "pause" if you don't. But remember we're all here to give constructive criticism and to help each other get better at this. Does anybody have any questions?

WL-C describes session time constraints and suggests that this will influence her behavior in the session. She brings up the physical rubric as a tool to help faculty remember the script, perhaps implying that they aren't responsible for having it fully memorized. WL-C rationalizes that focusing on implementation may "feel a little weird" to faculty—she lays out the rules but acknowledges they may be difficult for faculty to follow. She implies that it is natural for faculty to be drawn to discussing the content within this workshop and thus gives faculty a way to rationalize why there might be tension between these rules and what they instinctively want to do. She succinctly describes what will happen if faculty do not yell "pause" at appropriate times, and uses a plural pronoun that implies some shared ownership over potential implementation troubles across everyone in the session, including herself: "you should yell 'pause'...whenever we're not all doing the optimal job we can do."

5.4.2.3 Comparison of sessions B and C

There is significant overlap in the guidelines each WL lays out for faculty, but there are also subtle differences that contribute to WL-B's framing potentially cuing up more school-like ideas than WL-C's framing. In general, WL-B initially uses several school-oriented words and phrases ("welcome to class," doing your "homework," etc.) that might directly cue up faculty's ideas about acting as rule-abiding students in a traditional classroom, which WL-C does not. Both WLs acknowledge that faculty may find it difficult to follow the session rules, and both state what a repercussion would be (the WL "pausing" faculty). However, WL-C provides justification for why faculty might want to deviate from the rules, framing this as a natural inclination based on other aspects of the workshop, while WL-B instead focuses more on elaborating what it can look like for faculty not to follow the rules and describing potential negative repercussions of WL-B "pausing" faculty.

WL-B and WL-C also position themselves in different roles within the session.

WL-C uses phrases that imply that everyone is in this together, while WL-B makes deliberate moves to separate out faculty's responsibilities from her own. Both WLs mention time constraints in this session, but distribute responsibility for moving quickly differently: WL-B implies that faculty need to maintain a fast pace by elaborating on their "pause" immediately, putting more pressure on them to follow the rules closely (and adding more rules related to this), while WL-C implies that she needs to keep time, putting more of the burden on herself. Both WLs take up the framing of faculty being supportive critical friends for each other, but WL-B elaborates on this more extensively and in doing so implies that there is a correct way of implementing TPS that faculty should be striving for.

Thus, leading into these sessions, we already have a sense that these two WLs may make interestingly different facilitation choices, and that faculty might participate differently because of what ideas are being cued up for them and what they think they are expected to do.

5.4.3 First group implementation and critique

In these next two episodes, the first group of faculty in each session implements TPS, and are "paused" and critiqued on their implementation. Both groups present a question they wrote the previous day on the topic of simple harmonic motion. These episodes immediately follow the previous episodes. Because there is active faculty participation, we are able to look for evidence of faculty noticing and making sense of various implementation choices, and the extent to which they are given opportunities to do so. As such, we pay particular attention to who "pauses" the session (whether it is the WL or faculty), who provides justification for these "pauses," and the nature of these justifications. Again, we start with Session B.

5.4.3.1 Session B

WL-B: We have just had a great lecture on simple harmonic motion and now your instructors are ready to ask you their questions. When it's your time, get up to the front of the class. I'll give you my pointer…

Faculty presenter E indicates that she did not anticipate going first ("I thought we were [group] three"). WL-B explains how to use the pointer.

A (presenter): All right. So, I hope everybody enjoyed our fascinating lecture. [Advances the powerpoint slide to show the question and turns to face the screen.]
WL-B: Pause.
B: Fascinating.
WL-B: Does he need to say anything?
C: Yes.
WL-B: Not really. The lecture just happened. You can say "I've got a question" you don't need to go into anything about your lecture.
D: Because you introduced.
WL-B: Yes.
D: And you know it's simple harmonic.

The first "pause" from WL-B happens almost immediately after the first group starts. There is no evidence that faculty understand why WL-B paused before she explains, or that they take away transferable pedagogical implications afterwards. It seems unlikely that faculty would have paused the session this quickly themselves. Even after the "pause," faculty do not seem to understand WL-B's reasoning, quietly answering "yes" to a closed question for which the expected answer was "no."

From our perspective, there is little to no pedagogical reason for the "pause" based on what had occurred within the simulated implementation up to that point. This suggests that either her pedagogical justifications are very subtle and therefore also difficult for faculty to identify, or she is orienting to deviations from her exact vision for the session flow rather than reacting to a pedagogical misstep regarding TPS implementation. To the first point, her critique seems only weakly connected to faculty A's talk, because he did not "go into" anything about the lecture content and his body language did not indicate that he was about to elaborate further before he was "paused," and we are unsure why this would have been problematic.

To us, it seems likely that faculty A's talk did not represent an authentic part of what he would do in the classroom; instead, it seems more like a marker of transition into the mock implementation. Faculty A simply re-iterates the scene that WL-B had described. Whether or not WL-B was intentionally trying to draw attention to the session "rules" rather than a pedagogical move that could have implications for student learning, faculty also seem to interpret her "pause" in this way. Faculty D tries to re-articulate WL-B's reasoning for the "pause" by orienting to the specific rules within the session: "because you introduced" is an artifact of the workshop setting that would not translate well to the classroom. The second statement from faculty D, "And you know it's simple harmonic," could be slightly more pedagogically motivated, in the sense that it might build off WL-B's critique by implying that the physics content or topic would have been communicated already. However, this still seems disconnected from what actually occurred, since faculty A did not use any content-oriented language before he was "paused," which suggests that faculty A was focused on trying to understand WL-B's thinking rather than noticing and critiquing the presenters. Throughout this section, when faculty and WLs provide specific words or phrases when giving suggestions or critiquing the presenters, these words and phrases are indicated by quotations within the blocks of transcript.

A (presenter): So I have a question for you.
(WL-B turns off lights)
A (presenter): Oh, that's better.
A (presenter): So without talking to your neighbor…actually, do any of you need more time? Okay, without talking to your neighbor, please form your own opinion about what the answer is. Just take a minute for that.
E (presenter): Hold it to your chest so others can see.
A (presenter): Does anyone need more time? All right, go ahead and vote on the count of 3: 3, 2, 1, go.
(faculty vote)
A (presenter): Okay. Interesting. What I want you to do is turn to your neighbor and try to convince them that you're right. Just because you have the same answer, doesn't mean you're right or that you're done. Go.
WL-B: [after 2 seconds] Pause. You missed some key phrases.
F: "Explain your reasoning?"
WL-B: "Explain your reasoning."
G: Also, how much time we have.
WL-B: "You've got about…"
A (presenter): Oh, yes.
WL-B: …some period of time.
B: "About 30 seconds."
WL-B: And then what?
C: "Go?"
WL-B: "Go. Turn to your neighbor. Convince them that you're right." Say it with me,
WL-B and FP: "Turn to your neighbor. Convince them that you're right."
WL-B: "You've got about [FP: "30 seconds"] a min…30 seconds." Whatever. "Just because you both have the same answer doesn't meant that you're both correct, so make sure you explain your reasoning."
A (presenter): Okay.
WL-B: Then "You've got about a minute. [FP: "Go."] Go."

WL-B's "pause" and critique of "You missed some key phrases" lead into what seems like a school-like memorization game, where faculty are prompted to recite and listen to words and phrases from the prescribed script without any articulated pedagogical justifications for why these words and phrases would be consequential for student learning or engagement. We see evidence of this as WL-B takes the lead in the sequence of instructional moves while faculty contribute the words and short phrases that she appears to be looking for. Some of these turns of talk are well-characterized by an initiate-respond-evaluate (I-R-E) discursive pattern (e.g., "And then what?"-"Go?"-"Go."). WL-B also launches into reciting the whole script, prompting faculty to "say it with me," and saying it herself while faculty fade in and out of reciting and filling in the blanks.

We also notice that none of the "correct" phrases that faculty A said are acknowledged; instead, several of these phrases are elicited or re-stated by WL-B. For example, WL-B prompts faculty to state that the word "Go" should have been included even though faculty A said this initially, and when WL-B encourages faculty to recite the whole script with her, she starts off with previously stated phrases without indicating whether or not she noticed that faculty A had said these words. Through these moves, WL-B misses potential opportunities to praise faculty A's implementation and demonstrate that it is important to pay close attention to what faculty are doing well in addition to what they might be "missing."

Before the next exchange begins, faculty reason about the physics question in small groups, while the presenters walk around the room and WL-B voices some additional comments about implementation. We resume as the faculty presenters regain the attention of the whole room.

E (presenter): Okay. Let's vote again. 1, 2, 3, vote again.
A (presenter): Hm, so, let's see, we have a lot of (B's). This might be something…
WL-B: Pause. Pause. You are at a teachable moment right now. A (presenter): [*Laughs*] Go nuts.
WL-B: You're at a very teachable moment. [Another WL] and I anticipated

that this would be your result.
WL-B: Did you get the result that you wanted at the end of the second vote?
A (presenter): Absolutely not.
WL-B: Correct. There are some choices that you can make right now. Telling them about the distribution is not one of them.
A (presenter): Okay.
WL-B: Okay, so the best thing to do is to say, “Is there something about the question... is there a question about the question that you would like to ask? I didn’t get the result I was looking for.”
WL-B [to other faculty participants]: Go ahead. Are there any questions you would like to ask?
A (presenter):[*pointing to a faculty participant with their hand up*] Go ahead.

This is the final pause of this round of TPS implementation, again initiated by WL-B and done fairly quickly, leaving little opportunity for other faculty to have interjected instead. WL-B positions herself as more expert than faculty participants by suggesting that she and another WL anticipated this voting outcome, hinting that the faculty presenters did not anticipate this when they wrote the question. WL-B also takes an authoritative stance to redirecting the TPS implementation, while faculty A appears to be amenable to WL-B’s interjections and welcomes her critiques. Although WL-B states that faculty A has a choice in what to do next, she presents the “best” or correct path forward and dismisses faculty A’s approach as incorrect without justification. No space is made for alternative suggestions as WL-B leads faculty A into resuming his pseudo-instructor role with her suggested prompt.

Overall in this episode, we note that WL-B always “pauses” the session, which together with the introductory framing indicates that faculty who are acting as pseudo-students/critics did something wrong. WL-B often “pauses” the session at times when it would have been unlikely for faculty to do so—only the third “pause” might have provided an opportunity for faculty to interject first. “Pauses” are followed by closed discussion (WL-B prompting faculty to say certain phrases and evaluating their responses for correctness, faculty filling in one word or one phrase answers) or lecture that points to a single correct way of implementing, with little or no pedagogical justification. There are few if any detours from a memorization or recall game.

5.4.3.2 Session C

WL-C: Here’s the set-up. We’ve just had a great lecture on simple harmonic motion and now they’re ready to ask you their question.
N (presenter): Hopefully you guys can see this. Since we’ve explained to you everything there is to know about simple harmonic motion. We’re going to work on this short activity. Please take a few seconds to read the question and think about the answer, first on your own. Don’t discuss with the person next to you. I’ll give you a second to just read it. Think about it.
L (presenter): Do you need more time? Okay.
FP: Yes.
L (presenter): 10 more seconds.
N (presenter): Let’s get ready to vote. I hope you all have your four color voting cards. We’re going to vote on three, okay? One, two, three.
M (presenter): Yeah people. Yeah, this is a little bit painful, because as I can see we have less than (15%? 50%?) of the people choosing the right \\answer.]
WL-C: \\Pause!] Does anybody have anything to say?
FP: Interesting!
FP: Pause, pause.
FP: So you’re not supposed…
WL-C: When you have less than 80%, you don’t really want to tell them what the good or the bad is, you just want to move on to the getting them to talk.
M (presenter): Okay.

The WL-C initiates the first “pause” of the session. She invites faculty to contribute, which some begin to do, but WL-C then answers her own question by articulating the implementation “rules” without pedagogical justification. There is some evidence that faculty agree that a “pause” was appropriate, and also that they are orienting to a deviation from the implementation rules (“you’re not supposed (to)...”) to justify what should be changed. At this point it seems like the norms are school-like based on contributions from both WL-C and FP, but there is some evidence that WL-C and FP are on the same page about the “pause.”

M (presenter): And then uh, well, what I would like to do is encourage you, mistakes just happen, alright. I’d like to encourage you to talk to your colleagues and try to convince them about your answer and then you’re gonna go to another round of votes. To the count of three, one, two, three, and let’s go.
Multiple FP: [immediately] Pause.
N (presenter): Talk with your neighbors.
R: \\Give a time.]
S: “Talk to your neighbors, give your reasoning, blah, blah, blah.”
M (presenter): Sorry?
S: Oh I’m sorry.
E: We all have the same answer, we must be right.
WL-C: There was a pause there? What was the comment?
R: Give a length of time for us to discuss.
N (presenter): So say like “You have \\one minute”]...
FP: “You have 30 seconds”
N (presenter): “. . . to discuss with your . . . ”
R: “You have 30 seconds,” something like that, yeah.
N (presenter): Good suggestion.
L (presenter): Okay, 30 minute–, 30 seconds to discuss with your partner. Go!

Multiple faculty participants contribute to the second “pause” of the session. There is very little intervention or contribution from WL-C: she only speaks once as the session is paused and then resumed by the presenters. When WL-C does intervene, she invites a faculty participant to revoice his contribution so it can be heard by everyone, and thus helps faculty to narrow in on one suggestion that is picked up by two of the presenters to launch into a “redo.” Therefore there is significant ownership among the faculty presenters in whether or not to accept the suggestion and the duration of the critique. There is also a sense of teamwork among presenters—two faculty presenters step in to help their colleague who was "paused."

Turning to faculty-faculty interactions, many faculty contribute to fill in "gaps" in the script that were not said by the presenters. There is a shared sense from those who speak that saying all the prescribed words is desirable: they all make similar suggestions about what words to say, typically without elaboration or debate, and one of the presenters acknowledges value in these suggestions. Suggested repairs that were very close to what a presenter said ("talk to your colleagues" versus "talk to your neighbor,") and "blah blah blah" might demonstrate a lack of attention to what was communicated by the pseudo-instructors and a focus on the goal of saying the whole script. Notably, one faculty participant does provide justification for an implicit suggestion: she takes on the role of a pseudo-student, saying "we have the same answer, we must be right," which seems like a cue to add the phrase "just because you have the same answer, doesn't mean you are both right." In taking on this pseudo-student role, she is intentionally playing out a negative consequence of not including this phrase. Thus, there is evidence that at least some faculty can go beyond filling in memorized words, and that they perceive that providing justification could be useful or appropriate here. However, no one contests a lack of pedagogical justification when it is absent.

Before the next episode begins, faculty who are acting as pseudo-students discuss the physics in small groups while some of the presenters circulate the room. We resume as the presenters regain the attention of the other faculty.

O (presenter): Alright. Now that you've had some time to think about it and discuss with your neighbor, on three we're going to do another vote. Get ready to vote. One, two, three. All right. Much better. The correct answer was ...
Multiple FP: C.
O (presenter): Yes, exactly. That's because if a harmonic oscillator has the same mass and the same type of spring. They have the same ...
Multiple FP: Frequency, period, frequency.
O (presenter): Ah, right. Same frequency, same period. That's really the same thing. All right. Well great!
WL-C: Great job.
FP: Thank you [starts clapping, other faculty clap].
WL-C: That was great. Thank you for being the first volunteers. Any comments about the implementation when you're looking through these questions [motions looking at a rubric]?
G: I think the answer E is a little bit confusing. They will not cross.
WL-C: Okay, so we'll get to the question in a minute.
G: Oh, okay.
WL-C: No that's okay. It's because it's easy to go straight to the question, I want to make sure we're spending enough time on the implementation. [G: Okay.] So how they did the think-pair-share.
B: We already paused them, right?
WL-C: Yeah. You feel like you already said everything there was to say?
S: Could probably be a little bit faster, I'm not sure. The whole thing. Less time talking. That depends on the person who does it.
D: When we were encouraged to convince each other there was no statement made about having the same answer. So like our group we all had the same answer and we were kinda like "Yeah we're all right."
L (presenter): Ohh, that's right.
D: "Just because you have the same answer doesn't mean that you're right."
WL-C: There you go. "Turn to your neighbor, convince them that you're right just because you have the same answer..." [to M, who starts speaking quietly] Yes, you got it.
M (presenter): "You have the same answer, you're going to have a conversation and you will figure out if you change the answer... (inaudible)"
[another faculty member asks something inaudible (to us and potentially to WL-C), pointing to the question on the screen]
WL-C: Is this a comment about the question?
Other FP: Yeah.
FP: (We really want to comment)
WL-C: Okay, so here's why I'm working so much...it's very easy to talk about the question and the content and there are lot of resources out there for that, but this is the only time we'll all get feedback on the implementation. So we will get it, I promise. But you only get 2 minutes. I think that comment was about the getting students to engage in discourse even if they have the right answer. That was a very good point. And you guys did lots of stuff awesome.

So first group, I'm very impressed.
WL-C: Yes?
G: They could have maybe gone around to listen to our conversations.
WL-C: I noticed one person did. I didn't know if she was representing.
L (presenter): Yeah, I was, I was the person that kept time and went around to make sure that you guys were still talking.
G: I missed when you said "You have 10 seconds left." Maybe you did.
S: No there was that.
G: Sorry. Okay.
WL-C: I'll point out in this we're trying to really do it so when you say, when you turn around, you better not say 10 more seconds unless you would actually do it in the class. Just because we're trying to, that'll help us know whether or not you'd really do it.
WL-C: Okay. Now you have 2 minutes for the question itself.

When faculty finish the implementation round, the WL praises and thanks the pseudo-instructors before eliciting additional evaluation from participants. Faculty want to go straight to the content of the question when given the chance to provide comments about implementation, which conflicts with the "rules" WL-C outlined. However, WL-C is diplomatic about guiding faculty to comment on the implementation instead of the science content or the question itself. As in her initial framing, she expresses empathy for faculty's desire to talk about the content, framing this as a normal and natural response, and promises that they will talk about it later (their ideas are valuable, but it is not yet time for them to voice these ideas). She does indeed open up this space to discuss the question content at the end of this segment. When one faculty participant initially argues that there is not anything left to talk about with respect to implementation, WL-C invites faculty to judge whether or not there is more feedback left to offer ("You feel like you already said everything there was to say?"), and successfully initiates a faculty-led critique.

WL-C also uses several of her moves to give presenters agency to evaluate or correct their own instruction. She shares the responsibility of restating the words in the script with a faculty presenter by stopping herself to let him complete the lines when he begins to say it with her, passing off the task of saying the scripted words to him with an encouraging, "yes, you got it." She also defends the presenting team against a critique ("I noticed one person did, I don't know if she was representing") while still leaving room for discussion to continue—the presenters are given the agency to determine whether or not there was an error. Other faculty follow suit in giving presenters the benefit of the doubt during critiques. WL-C gives general praise to the presenters at multiple points during this discussion, acknowledging the value in critiques but also softening them with "you did a lot of stuff awesome."

In this session, faculty sometimes justify phrases in the TPS script based on likely student responses. These pedagogical justifications emerge without a direct prompt from WL-C, which shows us that faculty are able to do this given the opportunity, and we note that WL-C has created space for this to happen by giving a large amount of control over the discussion to faculty. In particular, faculty choose to revisit an idea from an earlier "pause" that was not well-explored at that time, that "there was no statement about having the correct answer." In both instances, faculty justify the importance of this statement by stating that from a student perspective, it is easy to assume they were all correct when they agreed on the answer. As in the earlier case, this faculty participant talks about himself in a student role in order to reason about these pedagogical implications. He then recounts the relevant line from the script. WL-C uses a later turn of talk to highlight this comment, generalizing that it was about "getting students to engage in discourse even if they have the right answer," praising this reasoning and de-emphasizing the exact words by not repeating them again.

Despite some pedagogical sense-making, however, many comments from faculty still do not showcase much pedagogical depth. Faculty take responsibility for critiquing the TPS implementation but typically do not justify extensively why the suggested change or modification, which was often closely aligned with the prescribed script, mattered for student learning or engagement. For example, when faculty mention the need to walk around the room listening to student conversations, which could have multiple pedagogical justifications (e.g., learning about the substance of students' thinking [Coffey et al., 2011]), the only justification that is voiced is about making sure participants were discussing at all. No one elaborates on this further. Later on, faculty G's comment of "I missed when you said we had 10 seconds left" seems rely on correctness as a justification, since no other justification is voiced. It is assumed that this was a phrase they were supposed to say. Faculty could be leaning on correctness to hasten conversations and/or because they are not easily able to articulate the underlying pedagogical justification. When there is limited or no justification, we cannot distinguish between these options.

5.4.3.3 Comparison of sessions B and C

While the first "pause" in TPS implementation and subsequent critique in each session share some similar characteristics, the WLs also use importantly different facilitation moves that seems to contribute to the development of different session norms. In particular, WL-C seems to be more successful at establishing a norm where faculty quickly become willing to take over this process themselves, likely through her choice at the start of this episode in combination with the initial framing discussed previously. In both of these episodes, the WL initiates the first "pause" and speaks more than any of the faculty during the first critique, and both the WLs and faculty participants initially orient more to the session and/or implementation "rules" rather than describing any potential drawbacks of these implementation moves for students. The first few words spoken by the presenters are also highly similar. However, the WLs make different choices about whether or not to "pause" after these first pseudo-instructor words are spoken: WL-B chooses to "pause" here while WL-C does not. In Session B, the justification for the "pause" seems somewhat opaque to us, and faculty seem to struggle to follow her reasoning. In Session C, the WL pauses at a later point where the reasoning seems clearer to us and to faculty. Although faculty in Session C do not have the opportunity to fully articulate their justifications before WL-C explains her own reasoning, she initially asks a potentially open-ended question ("Does anybody have anything to say?") and faculty seem ready to respond.

Our claim that faculty in Session C quickly become more ready and willing to "pause" their peers at the start of the session is supported by what happens soon after the initial framing and the first "pause." Faculty in Session C take up an opportunity to "pause" and critique the presenters themselves, indicating a growing ownership over this process, while faculty in Session B are presented with a parallel opportunity and do not, instead waiting for WL-B to "pause" the implementation herself. Specifically, in both sessions, the second "pause" occurs just after the faculty presenters have finished instructing the other faculty (acting as pseudo-students) to discuss the physics prompt in small groups, and both faculty presenters follow the script to similar extents—they both include some of the prescribed phrases and omit some pedagogically consequential directives (including "explain your reasoning"). Faculty in Session C say "pause" immediately, while in session B there is a couple seconds of silence and then WL-B "pauses" the session herself. Similar differences play out in the critique that follows: WL-C says very little and lets faculty discuss amongst themselves, only intervening to draw out one faculty participant's justification and thus focus the group on one idea, while WL-B maintains control of the "pause" and begins a closed discussion with faculty participants, where they only are given space to contribute short words and phrases that WL-B evaluates for correctness.

The structure of the two sessions diverge towards the end of these episodes: WL-B "pauses" the session again and pushes for a particular pedagogical move, while WL-C debriefs with all faculty afterwards. This is in part because faculty struggle more with the physics question in Session B, thus leading to an extended mock implementation round relative to Session C, where most faculty participants answer correctly on the second vote and the mock implementation naturally concludes. We do note that when faculty in Session C were given the opportunity to comment on the implementation of TPS at the end of this episode, multiple faculty spontaneously brought up aspects of implementation that might have been missing during their small group discussions. This suggests that at least some faculty were either trying to pay attention to the presenters while discussing physics content or were able to partially reconstruct how this played out, even though it was not apparent that they were paying attention to this at the time. In other words, some degree of attention to the physics content in-the-moment does not seem to preclude later retrospective analysis of the presenters' implementation.

Across both sessions, we notice that pointing out deviations from the prescribed script and filling in the "correct" words is often treated as sufficient during the critique process. This holds true for both faculty and WLs within these design structures. Faculty tend to orient to what is "supposed to" be said, and do not often elaborate on the pedagogical justifications behind these words or how they are importantly different from the presenters' words. However, we note that this does not imply that faculty are not capable of providing pedagogical justifications, i.e., considering how students might respond and why. Instead, this serves as evidence that within these sessions so far this kind of justification does not seem necessary to faculty based on how the WLs and other faculty respond, as there is little accountability among participants for doing so. We also note that there is initial evidence that at least some faculty can and will spontaneously bring up pedagogical justifications when given sufficient space, which we see at the end of Session C. WL-C revoices this faculty's contribution several turns later, which might suggest to faculty that this kind of contribution is valued.

5.4.4 Considering drawbacks and affordances of a scripted instructional routine

In the following episodes, a faculty member in each session spontaneously raises a concern about the scripted-ness of this instructional strategy, suggesting that there may be reasons to adapt or modify it, and the WLs and other faculty respond.

5.4.4.1 Session B

WL-B: Okay, I would like everyone to pull their rubric out if you don't already. I should have had that out. There's some talking... there's questions that you're supposed to be using as a reference for pausing. I just want you to remember that they're there.
H: Do you use these exact rubrics for every single question that...
WL-B: Absolutely.
H: You do?
WL-B: How do you think I know them so easily? I think they just come out of my mouth.
H: I find it hard that \\it's like a scripted...]
WL-B: \\Well, you've never used it before.] It is. It is, but it projects. It's been, we've field tested in our own classes trying many many many many many different phrases. And we're not saying you need to use... in fact, [a WL] said it many times yesterday. We're not saying you need to use these words at your home institution. You do whatever you want. We want you to have an experience using the words that we know absolutely motivate your students to have the conversation that you want efficiently, on board, doing it, and gettin' 'er done.
H: Because, it seems like there's a lot of good parts to it...
WL-B: Yes.
H: explaining yourselves and...
WL-B: There are.
H: ...key words, but scripting is difficult for everyone.
WL-B: It is. But I didn't... I mean, it evolved. I didn't know it, right? And there's all sorts of people who didn't... Again, I'm saying you don't have to remember that, but having good words to use and getting your students

accustomed to that experience... it's the same all the time. I'm reminded the same way all the time. I know what my expectations are all the time, is a good thing.
I: Do you have the script online somewhere or do we have to (inaudible).
WL-B: No, in fact, the implementation... there was a piece of paper that I handed out yesterday that had a link on it. It was blue. And it's on...
H: It's also on the...
WL-B: And it's on...
FP: The handout.
H: PhysPort website.
WL-B: It's on the PhysPort website and it's on the CAE website.

Faculty H's initial question of "Do you use these exact rubrics for every question..." might encompass multiple driving concerns, such as: "Are you more flexible in your own class?" (interest in learning more about what WL-B does in her class, for the purpose of imitating or evaluating it), "Are you telling me to follow this exact pattern every time?" (interest in diagnosing WL-B's take-home message for participants) and "Are there good reasons for changing this pattern based on student behaviors?" (interest in diagnosing potential student responses to this pedagogical strategy and anticipating reasons to adapt or modify the pattern because of these responses). The WL-B addresses the first two plausible concerns, but the third concern, which could involve more pedagogical sense-making, is not really explored. For the first potential concern, she unequivocally states that she does use these exact words in her class. To the second potential concern, she acknowledges that faculty have freedom to teach however they want, but implies that adaptation is a more risky endeavor than adoption, where adaptation may not yield responses from students that are equally favorable to adoption. She provides emotional support and encouragement for adoption by acknowledging that it can be challenging to learn to teach this way, but that faculty can learn to do it with practice. She also provides some pedagogical justification for why direct adoption could be more favorable than changing the script over time, because using the same words lends a sense of predictability or stability to students. However, this encouragement and pedagogical justification is absent with respect to faculty's potential adaptation or modification. For the third potential concern, WL-B argues that this exact script has been shown to always produce desirable student responses, and there is little elaboration of what these responses look like. WL-B refers to the words in the prescribed script as "good words to use" in an absolute sense, where the possibility that particular situations or student responses might justify adaptations is not acknowledged.

WL-B's bid to frame herself as a trustworthy authority seems to limit faculty H's opportunities to voice his opinions and hear others' rationales for when it might be reasonable to adapt TPS. In this instance, it seems like faculty H has decided that an adaptation of this strategy would be worthwhile for him, but it is not clear to us or WL-B if her argument for adoption has swayed his opinion. Faculty H appropriately points out a key piece of the script that has significant pedagogical value ("explain your reasoning"), but WL-B does not highlight this and thus misses an opportunity to ground potential adaptations of TPS with meaningful pedagogical anchors and demonstrate some alignment with faculty H's thinking. Her words give faculty permission to modify the script when they go home, but not here in this practice opportunity. We know that when faculty get home they will make these decisions for themselves with or without her official permission. Because WL-B would not be able to hold faculty accountable for following her rules in their own classrooms (since they do not have a sustained relationship outside of the workshop context), her "permission" seems somewhat superficial. Because WL-B does not invite faculty's responses or reactions, it is difficult to predict what they might do when they go home and why. The only faculty contributions that follow this exchange echo and further justify direct adoption.

5.4.4.2 Session C

A: So I have a meta question. Throughout the semester all the lectures will be like that, and I'm wondering if, will the students will get sick of it eventually if we just do this religiously, say the same thing each time. "You have 30 seconds," or "the fact that you have the same answer doesn't mean that you...," each time saying this might get old quickly I feel.

WL-C: I think if you're getting the sense that they know all of this very, very well [A: Mhm.] then making that shorter and figuring out what's the key part like, "Convince them you're right. Remember to..." you know, that "if you have the right answer you don't know, go." That's okay. Right now it's a script to get all the parts in [A: Yeah, sure.], but if it starts to be absurd to you it's probably absurd to them.

L: After the (third week) I don't think you have to say any of that I think just "Go!"

E It's comfort to them though.

L: Once in awhile.

N: Sometimes routine is good.

K: I think [another WL] would disagree. I think [WL] would say [WL-C: Do it every time?] that at least on the convince your neighbor part he always wants to build in that (inaudible) for that tension. That's what he was saying.

WL-C: I would say you still need to follow all of these questions so if they're not feeling a sense of urgency, if they're not feeling that they need to convince people then you are not saying the right thing.

FP: I sometimes underline, I often tell them, "If I don't hear arguing you guys aren't doing your job."

FP: I don't think [WL]'s style will necessarily work for everyone. Not everybody has exactly the same classroom presence and...

WL-C: \\And [WL] says that as well.]

FP: \\...the same kind of relationship with their students.] I think you can do all these things without inciting a riot. You know what I mean? Without feeling the need to give your students pitchforks.

WL-C: Right. There does need to be a sense of urgency to get them moving [FP: Yeah, definitely] but you don't have to be [WL] to do this job.

284 WL-C: Okay. Any questions in our two minutes about the question?

When faculty A brings up potential negative repercussions of using the same words every class ("students will get sick of it"), WL-C acknowledges the validity in his concern (that it might seem absurd to students) and tells him to trust his instructional intuitions about how students are likely reacting. She positions him as knowledgeable about how his instruction is coming across to students, how his students are reacting. Adapting the script in response to potential negative student perceptions is not shut down or discouraged; instead, WL-C encourages faculty to experiment and figure out "key parts" of the script as their students' needs shift.

After WL-C speaks, faculty also voice a variety of opinions in response to faculty A's question, which WL-C may have supported by stating "I think" instead of asserting certainty or authority. Another WL is strongly positioned as an epistemic authority in the conversation, first by faculty K and later by WL-C. Faculty K's appeal to the WL's authority seems to limit some pedagogical sense-making: "(WL) would say" is used to support his stance that one piece of the script is critical to say every time. This might imply faculty perceive this WL to be less flexible and more supportive of direct adoption (though there may be room to modify some parts of the script, this WL's authority might set which parts should be adapted). Some faculty argue for more adaptation/less strictly following this WL's script but often without articulating specifically what they would adapt, e.g., it is unclear what is meant by "(WL)'s style" or "all these things". The WL-C brings faculty back to focusing on markers of student engagement, but she also goes along with faculty's use of the WL as an epistemic authority to lend weight to a later argument ("(WL) says that as well").

5.4.4.3 Comparison of sessions B and C

In Session B, a concern about using the same script many times is voiced when WL-B draws faculty's attention to the TPS rubric for the first time. WL-B acknowledges that adaptation is possible, but implies that it is risky and might result in students not having desirable conversations or not starting these conversations efficiently. She instead focuses on encouraging faculty to persist in learning and trying out this exact script without any modification, making her the strongest proponent of adoption across all of the faculty and WL in these two implementations. She briefly describes ways that students would benefit from an instructional routine, but also leans on a WL and the other developers as epistemic authorities to defend the effectiveness of the prescribed script. This approach may conflict with faculty's current thinking about how an adapted version of this strategy would better match their instructional preferences, and a bid to identify key parts is not supported by WL-B.

In contrast, concerns about scripted routines emerge spontaneously in Session C when a faculty participant slightly interrupts the session flow to ask a "meta question." WL-C positions faculty as knowledgeable about how their students are responding and describes some intuitions that might lead faculty to shift their instruction. By doing so, she encourages faculty to thoughtfully adapt this strategy and communicates that she thinks they are capable of making these pedagogical judgments. Here, the identification of key parts of the strategy is brought up by WL-C and debated among faculty. Faculty hold the floor for more extended periods of time than in Session B and are given sufficient space to complete their sentences or thoughts without being interrupted by the WL. Faculty are typically the ones who bring up the exact steps in the TPS script and appeal to authoritative sources such as the developers' ideas, while WL-C spends more time articulating alignment with faculty's ideas and drawing attention to the general pedagogical foundation beneath what they are saying.

5.4.5 Summary

In these two sessions, we see evidence of faculty talking and acting in ways consistent with both "doing school" (primarily in Session B) and pedagogical sense-making (in Session C). In Session B, we see faculty only justifying pedagogical choices based on alignment with the prescribed strategy, filling in words and short phrases to respond to the workshop leader's prompts, waiting for the WL to initiate implementation critiques, and bringing up but not being given space to explore ideas that contrast with the WL's expert ideas. These behaviors are strongly reminiscent of "doing school": faculty seem to be going through the motions of enacting this strategy in order to meet the expectations of the workshop leader, we do not see strong evidence that faculty understand the purpose of the steps they are enacting, and the workshop leader aims for faculty to come to know predetermined answers to the exclusion of helping them develop sensible models for themselves. In contrast, in

Session C, we see faculty providing some pedagogical justifications for instructional moves based on how they think students would respond, speaking in complete sentences, developing ownership over the critiquing process, and debating alternative instructional moves with their peers. These behaviors are reminiscent of sense-making: faculty are beginning to critically analyze the prescribed implementation steps by providing evidence-based hypotheses to support their claims and debating with their peers using logic that is sensible to them, and session norms have been established where faculty spontaneously initiate these conversations themselves.

Throughout this subsection, we have looked at how variations in the workshop leader's actions and talk seem to have shaped the session in importantly different ways for faculty, leading to the observable differences we just described. These contrasting facilitation moves are summarized in Table 5.1. We note that an exclusive focus on RBIS adoption and "doing school" seem interestingly linked, as do considerations of potential RBIS adaptations and pedagogical sense-making. We also note that faculty enact some school-like norms in both sessions, such as inclination to focus on correctness without providing pedagogically–motivated justifications, which suggests that the structure of the session, the highly prescribed implementation script, and/or other experiences in the NFW may also cue up "school." While our analysis of Session C demonstrates that the session structure and the task itself could also be used as resources that support sense-making, this leaves us with questions about the rest of the NFW that we will continue to consider as we analyze other sessions. Additionally, while our analysis of session C shows initial evidence that faculty are able to provide valid pedagogical justifications, and sometimes do so without direct prompting, we do not see evidence of faculty extensively elaborating on their reasoning. Again, we will look for further evidence of faculty's pedagogical reasoning as we analyze later sessions.

Table 5.1: Workshop leader facilitation moves that might cue up or sustain school-like or pedagogical sense-making norms.

Facilitation moves promoting school-like norms	*Facilitation moves promoting sense-making*
Framing the session with explicitly school-like words	Encouraging faculty to use available resources to help them remember what to do
Implying that faculty are likely to break the session rules when setting and enforcing them	Acknowledging that following the session rules may "feel a little weird" to faculty
Emphasizing that faculty responsibilities in the session are different from their own responsibilities	Indicating shared ownership for following the session rules
Asking faculty to fill in implementation critiques at times when they are not likely to appear competent	Setting faculty up to be successful at analyzing instruction by choosing an accessible example and eliciting their responses
Treating statements of "correct" prescribed steps as sufficient critique of RBIS implementation	Hypothesizing about student behaviors that might justify a variety of instructional moves, and encouraging faculty to do the same
Infrequently praising faculty, primarily focusing on repairing deviations from prescribed words	Frequently praising and encouraging faculty
Asking exclusively closed questions and elaborating on faculty's short responses, sometimes interrupting them	Asking some open questions and allowing faculty to speak in complete sentences

5.5 Faculty's pseudo-student experiences

5.5.1 Overview

In the first half of this section, we will consider two cases in which groups of faculty act as pseudo-students to experience a task designed for physics or astronomy students. In both cases, workshop leaders act as the instructor, and we consider the interactions of the one focal group that we recorded. In the analysis below, we identify additional evidence of faculty adopting school-like framings within the NFW, we show evidence that faculty may experience collaboration troubles as they work with their peers in this school-like mode, and we show instances where these school-like frames (and collaboration troubles) are both reinforced (potentially cued up) and to some extent ameliorated by a workshop leader.

In the second half of this section, we consider how faculty reflect on these pseudo-student experiences and other collaboration troubles they have faced. In one case, these reflections happen spontaneously immediately following the activity and in a later session, while in another, the WL specifically asks faculty to reflect on what they just experienced. While each of these reflections is fairly limited in scope, e.g., faculty are not asked to elaborate on what they notice (by the workshop leaders or by their peers), we describe these instances in order to demonstrate that faculty are able to distill relevant takeaways from these pseudo-student experiences and could potentially engage in pedagogically rich discussions that build on their own observations and experiences.

5.5.2 Faculty acting as pseudo-students

5.5.2.1 Whiteboard work

This 45-minute session focuses on incorporating interactive engagement into upper–level physics classes. The part of the session we present detailed analysis of begins ~ 25 minutes into the session, and comprises the only time during the session when faculty work together in small groups. The WL primarily lectures for the first $\sim$ 15 minutes: she frames the session and describes her perceptions of new faculty's experiences learning to teach at the university level and related advice; she describes what physics students tend to struggle with in upper level physics classes, such as recalling knowledge from past courses, and how this might be broadly addressed through instruction; and she shows mathematically–advanced representations of physics phenomena that she suggests might be valuable for students to engage with in these classes. In the subsequent $\sim$ 10 minutes, the WL simulates RBIS that make use of embodied cognition (asking faculty to move around the room to represent different charge distributions) and small whiteboards that faculty write on individually (asking each faculty participant to write down what a dot product is). She intersperses this implementation with reflection by pointing out various aspects of her own implementation to faculty and encouraging them think about why these instructional moves might be beneficial for their students. Faculty also ask a few questions about instructional strategies and there is some brief discussion.

The analysis that follows, from this point up until 5.5.2.2, is adapted from our earlier work, [Olmstead and Turpen, 2015], where most text is taken verbatim from that publication.

Here, we consider the interactions of the four faculty members in our focal group, given the pseudonyms Ted, Maggie, Rachel, and Brad, as they work on an activity about conceptualizing plane waves. A large whiteboard and markers placed on the table between them and are needed to complete the task. 9 minutes of this session are allocated to this small group work, and we start our analysis at the beginning of this 9 minutes. We find that although these faculty appear genuinely immersed in enacting student roles, their behaviors do not exemplify cooperative, equitable, or intrinsically motivated student behaviors that we would want them to bring out in their students. Specifically, as we describe below, many of their interactions are consistent with markers of low coordination with their peers, and they focus more on "doing school" than sense-making.

As this small group work begins, all four faculty are seated and read instructions projected at the front of the room that will lead them through the activity. The first step directs them to draw a grid on their large whiteboard, with at least 7×7 *points spaced approximately 2 inches apart. Maggie reads aloud softly. After a few seconds, before others appear to be finished reading, Ted stands up, takes the cap off a marker and leans forward as if to draw on the whiteboard. He glances back at the instructions as the WL begins to speak.*

WL: Alright. \\You have one minute to get those dots up there. Make them as square as you can in one minute.]
Maggie: \\Before we draw, why don't we actually measure it?]
Ted: But I mean approximately two inches [motions as if to start drawing]
Rachel: Those are gonna be a centimeter right? [pushes a piece of paper onto the whiteboard, blocking Ted]
Ted: I think approx-, I mean
Maggie: Listen we wanna be accurate,
Ted: \\Okay.]
Rachel: \\No no but at least you have a straight line] [pushes the paper towards Ted]
Maggie: \\2.54cm.]
Rachel: You have a straight line.
Ted: You do it. [shrugs and pushes the paper back towards Rachel]

Maggie, Rachel, and Ted take up competing aspects of the WL's instructions: Maggie and Rachel attempt to be highly accurate, which aligns with an interpretation that the squareness of the grid is important, while Ted makes several bids to draw the grid "approximately" and repeatedly motions as if to start drawing, which aligns with the directive to draw the grid quickly, in "one minute." They do not offer any justification for their arguments, which may suggest that correctly interpreting the WL's rules takes priority over deciding what level of accuracy is appropriate for the task. The way that faculty physically and verbally negotiate who will draw on the whiteboard and how this drawing will be done puts the whiteboard in the center of their conflict. Treating a group artifact as contested "territory" is a marker of low coordination in group work, as are "conflicts of insistence" (their conflict does not build meaning), and violation of turn-taking norms (faculty repeatedly interrupt each other and talk simultaneously) [Barron, 2000].

Immediately after this first set of interactions, the WL comes over to their group, pushes the paper off the whiteboard, and starts drawing a grid on it. She does not question faculty about what they were doing previously and her actions functionally discard Rachel's approach. Ted vies for the WL's approval of his idea, claiming "That's what I was gonna do until... (pointing towards his group)." When she walks away, Ted asserts to his peers "I was about to do that very same thing until I got in trouble."

As Ted complains about getting in trouble, he, Brad, and Maggie begin drawing points on the whiteboard. Ted is the only group member who is standing and draws points across the whole whiteboard without pause; Maggie and Brad only draw points near the corners and edges of the grid and do so intermittently. Twice, Brad pulls away when he and Ted try to draw a point at the same location on the grid. The second time, Ted laughs and remarks, "How many physicists does it take to screw in a light bulb?," while Rachel, who is watching, jokes that the grid is made up of "drunken points." As Maggie finishes drawing and pulls away, Ted reaches across the table and adds two more points directly in front of her. Ted stands up and re-caps his marker, and the following dialogue begins.

Maggie: This is why men don't get to draw things. No that's 8. Oh that's 8. You should let the women draw it. Ted: [Laughs] \\Well it's a good thing I work on non-Euclidean geometry.]
Rachel: \\It's hung like men hang wallpaper]
[Maggie says something inaudible. She erases and redraws several points.]
Rachel: Although to be fair we shouldn't say things \\like this],
Brad: \\I tried.] [Smiles and shakes his head.]
Rachel: because if \\they said this is drawing like women drive then we'll get in trouble.]
Ted: \\Yeah. Oh my god.]

Consistent with our initial claim, Ted's statement about "getting in trouble" and his interaction with the WL imply a "doing school" mentality: Ted is trying to appease an authority figure and to establish himself as a "good student" set apart from his peers. This episode also reinforces our earlier claim that the group members compete for use of the whiteboard, treating it as territory. Maggie tries to renegotiate her role both within the task and relative to the whiteboard, using gender to position herself and Rachel as more competent at drawing the grid than Ted and Brad. From an equity standpoint, it seems consequential that Maggie promotes her own participation by assigning herself a drawing task, when secretarial roles are more often implicitly or explicitly assigned to female students than male students and can limit female students' access to learning during group work [Esmonde, 2009]. Her comments also suggest that gender plays a significant role in how she perceives their unequal participation in the task, and may be indicative of a larger underlying tension throughout these interactions. Still, this discursive move opens up a more active role for her within the task up to this point, without explicitly pointing out that Ted has been allocating most of that responsibility to himself. Ted, Rachel, and Brad all smile or laugh, reacting as if it was a joke, and Maggie successfully takes on the role that she made accessible to herself, thus temporarily gaining control of the whiteboard. Rachel sustains an earlier aspect of "doing school" by revoicing Ted's phrase about getting in trouble, now providing it as a potential risk of making comments about gender stereotypes. She seems to perceive that an aversion to breaking the rules of the classroom will be a valid motivator within her group, thus assuming they have a shared, school-like desire to win the favor of the WL.

After completing the construction of the grid, the group responds to the prompt: "For *every point on your grid…connect the points with equal values of $\boldsymbol{k} \cdot \boldsymbol{r}$," where the WL has introduced a different vector $\boldsymbol{k}$ to each group and $\boldsymbol{r}$ is the position vector from the origin. For this group, $\boldsymbol{k}$ is the vector with components 1 and 2, which makes the solution lines for which $x + 2y = c$, where c is a constant. Maggie states, "I don't understand what she's asking" and re-reads part of the instructions aloud. Ted begins to articulate some ideas that will go into the solution, such as "$\boldsymbol{k} \cdot \boldsymbol{r}$ is just an equation so it's just lines," but leaves many of his sentences unfinished. Just before the exchange below starts, the WL pauses all the participants and gives them additional guidance about what to do next.*

Ted: So this, so $\boldsymbol{k}$ is, $\boldsymbol{k}$ is (1, 2). Right? That's a vector. [writing across the whiteboard facing himself]
Maggie: Wait why are you? Why \\don't you just draw it $\hat{x} + 2\hat{y}$?] [also writing on the whiteboard, on the nearest corner to her]
Ted: \\Just, we're doing an x component. So yeah, so $\boldsymbol{k}$ dot x, $\boldsymbol{k}$ dot $\boldsymbol{r}$ is gonna be equal to $x + 2y$. Right? (Maggie nods.) You with me?
Brad: Yeah.

Lending weight to our emerging claims, even as faculty progress to a more challenging part of the task, low coordination persists. The whiteboard still seems to be perceived as territory, once again primarily controlled by Ted: he writes equations near the center of the board, upside down to everyone but himself. Maggie asks "why" and her tone indicates that she is frustrated. Violation of turn-taking norms continue and now are more consequential towards developing a shared understanding of the solution. Although Ted is likely aware that he and Maggie speak simultaneously, he does not attempt to repair this potential social misstep, nor does he acknowledge what Maggie said even though she suggests a viable alternative representation. Ted looks at Maggie frequently and seeks signs of confirmation that she is listening to him, but neither responds to her proposal nor articulates all parts of his thinking, discarding some of his own ideas without pause or explanation. In these ways, Ted launches into constructing the solution independently as the session continues.

While other groups may have fared better, it is clear that these faculty struggled to collaborate here, and that the workshop leader's facilitation moves seemed to reinforce faculty's sense that they were "doing school" as opposed to doing physics. In particular, we notice aspects of faculty's behavior that mirror authentic but problematic student interactions that might arise in classroom settings such as unequal participation, a lack of attentiveness to others' ideas, and a focus on a standard of achievement that carries little weight outside of school contexts (which particularly speaks to the traditional school-like nature of these interactions). In the following subsection, we will consider a second example of faculty acting as pseudo-students and the ways in which their group interactions are structured and supported that are similar and different to what we have just observed.

5.5.2.2 Astronomy Lecture-Tutorials

Here, we again consider faculty working in a small groups as pseudo-students while a workshop leader (or two workshop leaders, in this case) simulates the instructor's role. In this session, faculty engage with a Lecture-Tutorial for Introductory Astronomy, a RBIS developed for non-majors, "Astro 101" courses. At the start of this session and in a previous session, the lead WL simulated other RBIS such as TPS using astronomy content, described his own instructional context, and analyzed the RBIS implementation that faculty experienced in the session, pointing out pedagogically relevant aspects of his instructional decisions soon after they occurred. Continuing from the session on the previous day, at the start of this session, he narrows in on the astronomy content relevant for this Lecture-Tutorial, namely, Hubble's law and the expansion of the universe.

For readers who are less familiar with astronomy content, I will briefly describe the physical meaning behind Hubble plots in order to provide a stronger basis for understanding faculty's talk in the episode we are about to analyze. While faculty voice some common ideas that are not canonically correct, for conciseness, I will wait for faculty's ideas to emerge in the transcript and only describe the complete and correct reasoning here. (Readers who are already highly knowledgeable about astronomy can opt to skip ahead to the next paragraph.) A Hubble plot is a line or curve that represents the speeds at which galaxies appear to be moving away from us, as a consequence of the fact that the universe is expanding, versus each galaxy's distance away from us. If the universe were expanding at a constant rate, the Hubble plot would have a constant slope: specifically, it would be a straight line with a positive slope, starting at the origin. This can be understood by noting that if the universe were expanding in the same way everywhere (uniformly) for its entire history, a galaxy's apparent speed would be directly proportional to its distance away from us. For example, a galaxy twice as far away from us (compared to a closer galaxy) at some initial time must be twice as far away from us at a later time, and therefore must appear to be traveling twice as fast to get there. The slope of the Hubble plot—the difference in the apparent speeds of two galaxies divided by the difference their distances from us—yields the expansion rate. As a complicating factor for observers, because light travels at a constant speed, we are currently seeing farther away galaxies at earlier times in the history of the universe.

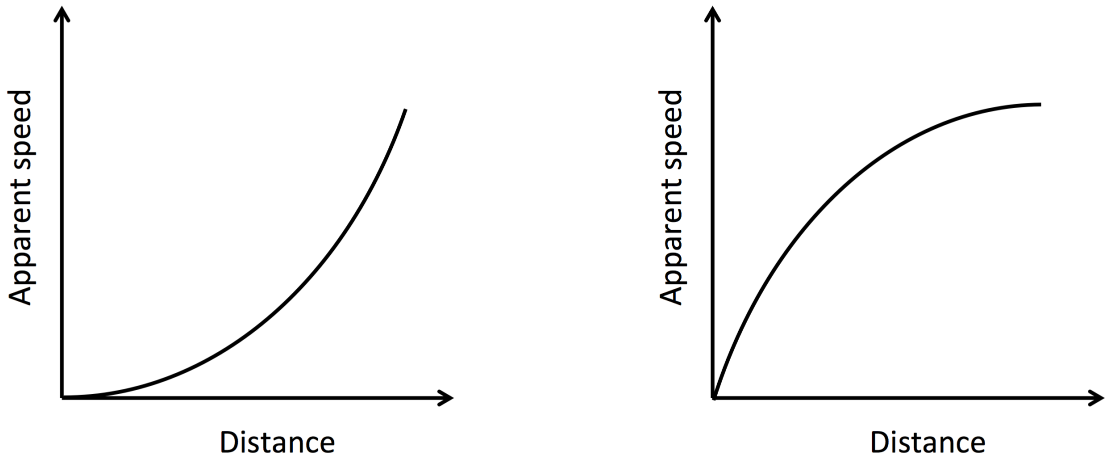

Figure 5.1: Hypothetical Hubble plots for a universe that is expanding at a decelerating rate (left) and a universe that is expanding at an accelerating rate (right).

This phenomenon is quite convenient for astronomers in a variety of ways, in this case, because looking at the apparent speeds of many galaxies at different distances (i.e., constructing a Hubble plot) allows us to see how the universe's expansion rate has changed over long timescales. Because of this, a Hubble plot does not need to be a straight line; instead, the Hubble plot for a universe with a changing expansion rate would be a curve, as shown in Figure 5.1. In this general case, the instantaneous slope at any given point on a Hubble curve/line indicates the universe's expansion rate at the point in the history of the universe when the light from that galaxy started to travel to us.

In this session, faculty are asked to start midway through a Lecture-Tutorial on Hubble's law. At this point in the Tutorial, two hypothetical Hubble plots for universes with non-constant expansion rates (and thus non-constant slopes) are introduced—one that would represent a universe that is expanding at an increasing rate and one that would represent a universe that is expanding at a decreasing rate—and students are asked to choose which one represents a universe with an accelerating expansion rate. A series of short prompts on the following page steps students through logic that would support them in choosing and justifying the correct choice. Before faculty begin to work, the WL emphasizes the importance of RBIS implementation and tells them that they will discuss the implementation of the Lecture-Tutorial (from the instructor's point of view) at the end of the session. He also gives them a set of "rules" for how they should act during this activity, instructing them to collaborate as if they are students in his class. ("Talk to your neighbor. Don't even start writing down an answer until you guys have come to a consensus. You must agree on your answer...This is your textbook...") From here, faculty begin to work.

The initial interactions among our focal group—given the pseudonyms Jade, Kurt, and Leah—are shown below.

Kurt: Okay then. "From the blank graph...draw a Hubble plot...for which the expansion rate increases throughout the lifetime of the universe."
Jade: So it's just a line? [Laughs] I don't know anything about this.
Kurt: I don't either, but I memorized it from the last session.
Jade: This is my only day I've been here.
Kurt: It's supposed to be like that [motions a curve that looks similar to the canonical answer].
Leah: The Hubble's law?
Kurt: It's accelerating expansion rate, right? The expansion rate increases throughout the lifetime of the universe, so it's not going to be a line.
Jade: The velocity is going to increase.
Kurt: It's either, no thinking sloppily, but it's either going to be this or that, right?
Jade: Is the acc-
Leah: This is (unsolved) question.
Kurt: No, given that...
Leah: Given that it's
Kurt: It's increasing
Leah: Increasing.

Kurt: Then draw this plot.
Leah: Are they, and the velocity is constant? Or it's like, a function? [motions a curve with a flat top]
Jade: The expansion rate...
Kurt: It's going to be like that, yeah.
Leah: It's flat?
Kurt: No.
Jade: [Looking at Kurt] Is the expansion rate the velocity?
Kurt: Expansion rate is the velocity of the thing growing. The velocity of the radius basically, yeah.
Jade: So it's definitely
Kurt: (inaudible).
Jade: So it's increasing but not a constant.
Kurt: Yeah, that's the point, I think.
[WL-1 is hovering nearby, looks over at what faculty are doing]
Kurt: Why do we have 2 plots?
WL-1: Because there's more than one question.
Kurt: Oh.
WL-1: Make sure again, that you all agree. You're drawing things, she isn't drawing things, you don't have one. Let me get you one. [Goes to look for another copy of the Tutorial, Jade thanks him.]
Leah: This is which question here?
Kurt: I thought we were doing 14, but I only see one question?
Jade: He told us to start with 14.
Kurt: Yes.
Jade: We're on Figure 6.
Kurt: Oh, sorry. Right here? Well...
Jade: But wouldn't, okay so without knowing anything about Hubble plots, so just the fact that the expansion rate is increasing...
Leah: But if rate is increasing it should be this [motions a line where the slope increases to the right; similar to the left-hand plot in Figure 5.1].
Jade: But if you drew this, if you drew that...
Kurt: Yeah.
Jade: Then all those rates are still increasing right?
Leah: [looking down at her copy of the Tutorial] Ah this is velocity.
Kurt: [to Jade] No, so this would be...
Jade: Right? Velocity is increasing.
Kurt: Expansion rate, so this straight one would be constant expansion rates.
Jade: Okay, but, that's my question here, is the velocity the same thing as expansion rate?
Kurt: Yeah, I think so.
Jade: So velocity is increasing this whole time. Even if it's straight, even if it's curved.
Kurt: This is not time though, this is distance.
Jade: Okay.

96 Kurt: So you're looking far away,
97 Jade: Distance...
98 Kurt: and looking at the velocity of something far away.
99 Jade: Okay.

As we begin our analysis of one focal group, we note that in the NFW, many participants are physics faculty and unfamiliar with astronomy concepts. This was the case in our data: it is clearly evident that none of the three faculty in our focal group are astronomy experts. At the start of their small group interactions, Jade states that she missed the previous day of the workshop and "I don't know anything about this," and Kurt responds "I don't either but I memorized it from the last session. It's supposed to be like that," and points to one of the two plots. Nonetheless, even though faculty could simply take up the "memorized" correct answer, we see some markers of sense-making in their interactions around this first prompt.

Jade asks questions like "is the velocity the same thing as expansion rate?," and shows persistence in building sensible ideas around this task. We note Jade's question is conceptually rich and highly relevant, as velocity refers to the apparent speeds of other galaxies as seen by an observer in our galaxy, while the expansion rate is a characteristic of the whole universe, and the difference between these is critical to understanding the plot. We also see some markers of high coordination in the interactions between Kurt and Jade, where they show attention to each other's thinking, such as repeating and elaborating on each other's words and respecting each other's turns of talk instead of speaking over each other. In these coordinated interactions, Jade and Kurt also hone in on the meaning on the x-axis, and start to unpack how it tells them about both time into the past and distance, which is also critical to their making sense of the Hubble plot. At the same time, we note that the collaboration among the group is not fully coordinated; instead, we see Jade and Kurt following one, conceptually–oriented line of reasoning together while Leah follows another, more mathematically–oriented one, interacting with Kurt to solicit his input (and interacting minimally with Jade).

Two workshop leaders circulate the room as faculty work in their small groups, and occasionally interact with our focal group, as we saw briefly in the transcript above. We will call the primary workshop leader, i.e., the one who facilitates the whole group interactions in this session, WL-1, and the secondary facilitator WL-2. In an extended interaction with WL-1, we see evidence that faculty are orienting to the WLs in a school-like way. Even though WL-1 asks a series of open questions of "Do you guys need any help, do you need any questions answered, is everything good? Let me know if you need any help," Kurt asks "What's the answer?" instead of articulating the conceptual sticking points that were emerging in their conversation. We see WL-1 model the kind of pedagogy we imagine he would use within his own classroom. For instance, he immediately solicits input from all of the faculty, not just Kurt; he asks a series of closed questions to guide them to the correct answer; and he quickly exits the conversation by walking away when it seems that faculty are on the right track, leaving them to complete the line of reasoning and come to a conclusion with each other instead of relying on his authority. He also moves Leah away from the jargon-rich question that she asks initially, such that she ultimately asks the same conceptual question that Jade had voiced earlier ("Is expansion rate the same thing as velocity?"). Following WL-1's intervention, faculty (particularly Jade) seem satisfied with the response that Kurt had initially proposed, and move on.

Immediately following the interaction described above, Kurt and Jade talk quietly and appear to examine the next prompt together, pointing at their papers, while Leah works independently. This continues for a few minutes, until WL-2 comes over and interacts with their group as follows.

WL-2: I'm a little bit concerned that y'all are supposed to be collaborating. She's just started to write her answer down to the question at the top of this page and she's flipping back to the previous page and you two are talking about the next question and getting ready to write things down so, I don't know how that coming to a consensus through your collaboration, and writing your answers down together before you move on, is happening.
Jade: Okay.
WL-2: Try to fix that.
Kurt: What do you think?
Jade: Which one are you on?
WL-1: [to the whole room] Hey by the way, by a show of hands, how many of you are on the last page or done? Last page or done? Okay, we just got a couple of more minutes.
Kurt: She's doing calculations.
Leah: I don't know, I just, it looks like it's proportional to square root of distance, I was just able to figure out if it'll go with time.
Kurt: Oh I see.
Leah: It's asking about the time difference right? What do you think?
Kurt: We're sort of thinking qualitatively.
Jade: Yeah. Just kinda like, I'm just using slope for a lot of this stuff, so like if it's a constant slope or it's a slowing down type of slope or ... I think that's the path he was leading us on.
Kurt: I'm ...
Leah: It was constant, it would be proportional to, a little bit distance, right?
Jade: So yeah it would be a straight line.
Kurt: If it goes that way, how you ... find ...
Jade: It'll be a straight line if it was constant.
Kurt: Yes.
Leah: Versus time.
Kurt: I understand what you're trying to do, but ...
Jade: Oh you're doing versus time.

Kurt: She's trying to convert this thing into a velocity (inaudible) time thing right? But that's not that. This is a bunch of things. It's not a single particle going. Alright, this is not v as a function of x and x as a function of t, it's not like that. This is a different picture, so you have multiple points and you're here. You're looking at here, this is x and this guy over here has a velocity. Maybe it might help to think of it, this is how I'm thinking, maybe not the best way to think but, I'm thinking, okay so let's pretend that the universe was accelerating, the universe was expanding at a constant rate for a very long time, but then suddenly accelerated at the last second, what would I see then? If I already have a universe like that, and if it just expanded very much in the last minute, the guy, the guy that's nearest to me, would be, I would immediately see it but I just, going very fast suddenly. This guy, I think I'm seeing an old version of this guy. I'm still seeing this guy to be moving with the old rate.
Jade: Cuz he's further away.
Kurt: Yes.
Leah: Yes.
Kurt: That's my reasoning.
[WL-1 calls time.]

At the start of this episode, WL-2's intervention has some school-like qualities similar to what we saw in the TPS session. Her statement that faculty are "supposed to be collaborating" orients faculty to the session rules rather than a pedagogical justification for why it might be desirable to shift the form of their collaboration, and her later statement of "try to fix that" implies that faculty's collaboration was somehow incorrect prior to this intervention. She very explicitly states what she noticed as problematic that caused her to come over to their group, and her statement of "I don't know how..." preceding a statement of what faculty are supposed to be doing in the session seems to be sarcastic, since she seems to imply that she does not think faculty are following these rules (and appears fairly certain that this is the case).

At the same time, WL-2's intervention does support faculty's collaboration and sense-making. We can see two aspects of WL-2's facilitation that contributed to the success of this intervention: first, although WL-2 emphasizes following the session rules, these rules illustrate how collaborative group interactions might be productively structured and thus drawing attention to them had potential benefits, and second, the WL had correctly identified markers of low coordination in their work (faculty pursuing independent solution paths [Barron, 2000]), and thus her intervention was timely and relevant. When the WL walks away, Jade and Kurt do take up her directive to re-initiate interactions with and get on the same page (literally and metaphorically) with Leah, immediately asking her "What do you think?" and "Which one are you on?" This leads them to realize that they were pursuing the astronomy task quite differently from each other and to start to discuss how to proceed together.

We notice some faculty talk moves following WL-2's intervention that seem to support development of shared understanding and sense-making. Once Jade and Kurt learn that Leah is taking a mathematically–oriented approach to the Astro 101 task, including "doing calculations" and trying to articulate a non-linear equation based on what a sketched plot in the Lecture-Tutorial "looks like," they call attention to their own approach, characterizing their own thinking as qualitative and illustrating the kinds of reasoning they might use. In particular, we note that Jade's use of the phrase "a slowing down type of slope" invokes ideas about a physical scenario one might infer from a non-constant slope on a Hubble plot using everyday language that might be more accessible to Astro 101 students. Jade's language clearly contrasts with Leah's language, and putting these options on the table may help faculty (particularly Leah) begin to weigh the affordances and drawbacks of these different approaches. We further note that Kurt does not ignore Leah's approach even though he disagrees with it; instead, he takes the time to articulate the logic that he thinks Leah is following and provide some justification for why it is flawed before launching into a longer explanation of logic that might support canonically correct, conceptual reasoning around this task. He also hedges his explanation by framing it as "how I'm thinking" (as opposed to how you should think) and "maybe not the best way to think," which might open up space for other faculty to disagree or state that they do not understand [Conlin, 2012]. While there are few turns of talk from Leah and Jade following Kurt's long explanation, and thus little evidence to suggest the extent to which they understand, these markers of high coordination suggest that their talk is structured to support sense-making.

5.5.3 Faculty's reflection on pseudo-student experiences

The previous episodes make us interested in how faculty perceive these pseudo-student experiences. In particular, we are less interested in whether faculty can collaborate on disciplinary tasks, and more interested in whether they might be learning about how to facilitate productive group work from these experiences. Thus, we revisit each of these groups of faculty and consider how able they are to reflect.

5.5.3.1 Reflection after the whiteboard task

We find two instances where faculty from the whiteboard task make bids to reflect together: one immediately following the completion of the task, and one in a later session where some of the same faculty (Maggie, Ted, and Brad) appear together in another focal group.

Faculty have finished the task and start to joke about the quality of their whiteboard drawing (suggesting that part of the table is gravitationally lensed).

Maggie: She's trying to get across the point of what exactly?

Ted: Just get us to understand– well we're working together, we're building teamwork skills, learning what a dot product...

Maggie: But this idea of the dot product. She drew the vector $\boldsymbol{k}$

Ted: The idea that the– yeah

Maggie: and the reasoning we had to go behind it was we needed to determine, using, what you had done over here the $\boldsymbol{r}$ dot $\boldsymbol{k}$, right? we had to figure out some type of formula for our line

Ted: Yeah.

Maggie: and then we took that and we said we know that it has to be what, parallel to it, right?

Ted: That $\boldsymbol{k}$ dot $\boldsymbol{r}$ equals a constant. The same value of $\boldsymbol{k}$ dot $\boldsymbol{r}$. That's an equation, we're in 2 dimensions, so what equation in two dimensions is a line. Or is a curve and it's linear in (x, y). And so we solved that equation, we found the set of different things for different values, for different constant values. And now we understand that not only are all these lines parallel but they're perpendicular to $\boldsymbol{k}$. Cuz they all have the same component parallel to $\boldsymbol{k}$.

Rachel (quietly): (It's contours in a contour map. It's beautiful. All perpendicular to the gradient.)

WL: Pause

We find it interesting that Maggie spontaneously encourages her group to reflect on the pedagogical intent of the WL once they have finished the assigned whiteboard activity. We also find it interesting that Ted initially orients to her question by providing very general suggestions about what students might be learning from participating in these kinds of small group activities. While we see potential value in reflecting on what students are learning about collaboration through interactive instruction, Ted's suggestions seem disconnected from his experiences in this session, where from our perspective, their group interactions were not highly collaborative.

It is unclear if Ted means to imply that their collaboration in this session was highly cohesive, but it is plausible to us that he is simply making broad statements that seem to align with what he thinks they were supposed to learn about the value of group work. In either case, he does not substantiate these ideas, as Maggie takes up his last bid to unpack the content learning goals for this specific activity, which seems to be her original intent ("But this idea of the dot product").

We are also able to revisit Ted and Maggie's thinking and interactions later on in the workshop. At the start of a session about collaborative group problem solving, another workshop leader asks faculty to choose which two topics (out of a list) their group would like him to focus on during the session. Maggie, Ted, Brad, and two other faculty participants are in our focal group during this session. During their conversation, Ted suggests "encouraging productive work in groups" as a topic he wants to talk about, with the justification that "sometimes I'm worried somebody would be like the, somebody would be, would take over the group and the other person, there will be like people that don't participate as fully." After other faculty veer towards a discussion of how to encourage students to be on task, and whether or not students should be forced to participate based on which they find better represents and prepares students for "real life," Ted says: "I find, I dunno I sometimes find it hard to work in groups, you haven't found it challenging, frustrating sometimes?" This utterance both expresses empathy for students based on his own struggles, and re-introduces his original concern into the conversation. After another faculty member confirms that she thinks everyone finds group work frustrating at times, he goes on to elaborate "Yeah I mean sometimes you find yourself arguing over things that aren't really working, different people move at different paces, and it can be certainly a challenge even for people that are motivated or you know who are (persistent)." However, no potential solutions for facilitating productive group work are articulated as Maggie keeps the group "on task" and makes a successful bid to identify a second topic that they want the workshop leader to discuss. (It is unclear to what extent the simplistic nature of the group task is limiting this interaction, but we note that it does not disrupt the pattern of interactions that preceded this for faculty to move away from potentially rich conversations without offering potential solutions, which seems to be linked to or justified by an inclination to adhere to the session rules.)

Here, we find it compelling that Ted spontaneously identifies a specific challenge that students may encounter during group work—one student "taking over" the group such that others "don't participate as fully"—that closely relates to what we identified as problematic in his earlier interactions. It is also interesting that Ted suggests possible emotional responses ("frustration") and challenges in coordinating thinking across many people ("different people move at different paces") that might be negatively influencing students' participation, and relates this to his own struggles with group work. We wonder if Ted's own struggles with group work and his desire to unpack these struggles could help faculty to identify markers of problematic interactions experienced within the workshop. However, the structure of this small group task discourages faculty from unpacking these genuinely difficult and potentially uncomfortable challenges together.

5.5.3.2 Reflection after the Lecture-Tutorial

Turning to our other focal group, soon after wrapping up the astronomy Lecture-Tutorial small group activity, the WL asks faculty to reflect on the "non-productive things" and "productive things" that occurred when they interacted with each other and with him during the Tutorial. He introduces this activity by explicitly trying to cue up the same kinds of conversation norms that they established in an earlier session where they were analyzing video of physics students and a TA (which we will revisit ourselves in the next section). This results in a fairly roundabout launch of this pedagogical analysis that may have been mildly confusing to faculty (e.g., "Imagine you just saw a video of your group and my interaction with your group. Except your group"). However, the conversations that follow suggest that at least most of the faculty participants got the gist of what he was asking them to do. Faculty are given a minute to discuss in their small groups, and the following exchange occurs in our focal group:

Jade: The first thing he asked us [was] if we had questions and I don't think we really even knew what we were doing. But he helped us get there, he never told us an answer or anything like that he just kept asking us questions. They also made sure we were all working together. We had an issue with that. Okay so those are important things to make sure you do. The purpose actually is cooperative.

Kurt: He didn't say that though. He sort of briefed the physics I guess, step by step.

Jade: Yeah.

This exchange is very brief and ends when Kurt jokes with WL-1 as he walks by their group. WL-1 makes a slightly early bid to end the small group conversations immediately after this exchange, and our focal group's conversation ends as they seem to wait for the whole group discussion to begin. Still, we can identify several interesting aspects of Jade's talk. First, she is able to identify pedagogically salient features of their interactions with WL-1 that are similar to the key points we noticed and summarized in their earlier interactions, such as "he never told us an answer," and "he kept asking us questions." In her talk, she in some ways uses their concrete experiences as a launching point towards more general education research principles (like the importance of guiding instead of telling), as others have suggested may be a characteristic of productively reflecting on instruction together [Horn and Little, 2010, Aubusson et al., 2010]. Jade also describes these elements as "important things to make sure you do," suggesting that she can see value in following these guidelines when facilitating group work in her own classroom. Moreover, Jade calls attention to their collaboration troubles and WL-2's intervention that we analyzed, recognizing that they "had an issue with" collaborating and that this intervention helped, cleanly articulating that "the purpose is actually cooperative." Again, this is a valuable guideline, which could apply to implementing many different RBIS, that she distills from this concrete experience. (We note that Kurt's disagreement with Jade may stem from a different interpretation of WL-1's instructions as opposed to a substantive disagreement about how the same event played out. Specifically, WL-1 only states that they should analyze his interactions with their group, which Kurt references, while Jade reasonably infers that their interactions with either WL should be considered in this conversation.)

Following this short small group discussion, WL-1 asks faculty to share their ideas and insights, which gives us (and participants in the session) the opportunity to hear some of what faculty in other groups have discussed.

WL-1: Okay, time. I'm sorry you guys that [things are timed so that I compressed] so badly. Comments? Questions? I would like to hear a little more of conversation about your thoughts and insights of what happened? Yeah.
H: We didn't see you.
G: Yeah, you ignored us.
WL-1: Is that good or bad?
I: Good, because we were doing the lesson.
WL-1: Right? People always ask me, do you jump, do you talk to every group, do you infiltrate the groups you go on. In the video you saw, [the TA] slid in and kind of sat there and, sort of not intrusive and, say, "Hey I'm here if you need anything." I have 400 groups.

Here, WL-1 begins by inviting faculty to share their thoughts and "insights," which suggests to faculty from the onset that he expects them to have valuable ideas. When faculty H and G point out something they noticed about his interactions (or lack of interactions) with their group, he asks a potentially neutral, open question "Is that good or bad?" that encourages faculty to evaluate the pedagogical affordances or drawbacks of his instructional moves for themselves. We note that despite what is voiced in this moment, this question may not have been perceived as entirely neutral, since he has portrayed himself as modeling expert pedagogy (and confirms this interpretation in the turn that follows). Still, Faculty I demonstrates that faculty are able to evaluate or at least rationalize the WL's behavior based on their recollection of their pseudo-student behaviors. WL-1 supports their interpretation by strongly agreeing with it ("Right?") and elaborating that this justification underlies parts of his pedagogical approach in class, potentially bolstering faculty's confidence by positioning them as able to discern something that others often miss. The WL also draws attention to the influence of class size on his pedagogical decision-making by contrasting this concrete example with the concrete example in the video, but does not dismiss the validity/worth of the pedagogical approach faculty saw in the video for small class sizes. By doing this, he suggests a need for faculty to flexibly adapt RBIS based on their class size (or more specifically their instructor-student ratio) instead of prescribing a single approach (his approach) for all situations.

While this conversation soon shifts to faculty asking questions about WL-1's instructional approach and WL-1 answering through lecture, some additional turns of talk from faculty provide additional evidence that they are able to notice and interpret events that occurred as they acted as pseudo-students. In particular, another faculty member points out "You did like a drive by, 'You guys are doing a good job, great'," which the WL elaborates on in a similar way to the above quote. We also see evidence that the faculty participant finds value in this approach, as she interjects "it was just a quick comment" and "yeah it worked." Another faculty member asks "Do you listen to what we're saying?," which could be pedagogically valuable unpack, since listening to students' conversations can allow instructors to better understand what their students are thinking. Again, WL-1 strongly agrees ("Absolutely") and provides a pedagogical justification for listening to students that aligns with our previous statement. In this case, however, the faculty participant is not encouraged to elaborate on their own reasoning, and this leads into a more closed question and answer period. Nonetheless, we consider these interactions to be signposts of the kinds of conversations faculty could engage in if given more space to reflect on pseudo-student experiences where the WL models potentially desirable pedagogical approaches.

5.5.4 Summary

These two sets of pseudo-student experiences were differently generative for faculty's reflection: in one case (the whiteboard task) faculty seem unable to identify useful pedagogical takeaways surrounding facilitation of group work, and in the other case (the Lecture-Tutorial), they do. We can begin to understand these differences by looking at the extent to which faculty were "doing school" versus sense-making around science ideas when working on student tasks together, which we summarize below. We will then return to the variations in structural elements (the WLs' actions and talk) that seemed to contribute to these contrasting pseudo-student/reflective sequences.

Before we begin to summarize and compare these sequences, we note that variations in faculty's interactions between the two groups are likely also contingent on differences in the characteristics of the faculty who comprise these groups, not just the WLs' implementation moves. While this is also a relevant caveat for the analysis in Section 5.4, there we considered larger groups of faculty and were able to identify ways in which faculty's tenancies seemed highly similar across the two sessions, which enabled us to make fairly strong claims. In this case, differences between these two small groups could be more pronounced; thus, our claims about how the WLs' different facilitation moves were consequential for producing the observed differences are more preliminary due to these limitations of our data. Nonetheless, we consider it useful to articulate plausible causal stories based on the other structural differences which we can observe.

When faculty engage in the whiteboard task, we see that many of their behaviors are consistent with Barron [2000]'s markers of low coordination in group work, which inhibits their ability to develop shared scientific (or in this case mathematical) meaning. We summarize faculty's behaviors alongside Barron [2000]'s markers in Table 5.2. Beyond this lack of coordination, faculty's talk and actions suggest that following the rules and winning the approval of the WL are important, but they do not justify why enacting these behaviors would be useful for any other reason. In these ways, we see faculty's pseudo-student behaviors as consistent with "doing school." Following this task, some faculty spontaneously make bids to reflect and/or discuss how they might ameliorate their students' collaboration troubles with their peers, but these conversations do not generate potential solutions for how to address these challenges when facilitating small group interactions. Faculty's reflective conversations immediately following this pseudo-student task focus only on the disciplinary content and faculty's comments about collaboration are vague and disconnected from their experiences. In a later session, one faculty participant's bid to unpack his experiences surrounding the same kinds of challenges that we identified in our analysis are not explicitly linked to this workshop experience and are not taken up by his peers. Thus, we do not see evidence of faculty learning to better facilitate small group work by reflecting on their engagement in this whiteboard task.

Table 5.2: Markers of low and high coordination in group work recreated from Barron [2000] Table 2, with some extended descriptions of how these markers were enacted in faculty's interactions.

Markers of low coordination in faculty interactions	*Markers of high coordination in faculty interactions*
Independent solution paths—Faculty do not respond to each other's ideas and/or do not talk to each other	Co-construction of solutions—Faculty hedge their explanations, leaving space for others to disagree
Reference to own ideas	Reference to other's ideas
Artifacts as territory—Faculty treat the whiteboard as contested territory	Artifacts as center of coordination—Faculty point to the same Lecture-Tutorial page as they build shared understanding
Individual monitoring	Joint monitoring of solution
Conflicts of insistence—Faculty do not fully articulate their scientific reasoning	Productive conflicts—Faculty interrogate alternative approaches
No response to contributions—Faculty discard or ignore some of their peers' ideas without pause	Transactional responses—Faculty repeat and elaborate on each other's ideas
Turn-taking norms violated—Faculty interrupt each other and talk simultaneously	Turn-taking norms respected—Faculty speak in complete sentences

During the Lecture-Tutorial, while some faculty in our focal group collaborate well from the onset, as a whole, they struggle to build scientific meaning together initially. These collaboration troubles are aligned with Barron [2000]'s markers of low coordination: faculty take different solution paths, some conceptually–focused and some not, and progress through the task at different rates. Some faculty again orient to the WLs as if trying to form alliances or play a school game, for instance, immediately asking "What's the right answer?" and using advanced scientific terms that they seem not to fully understand when interacting with a WL. All of these behaviors seem consistent with faculty "doing school." However, unlike in the previous case, faculty's interactions become increasingly collaborative throughout the session, as represented by the markers of high coordination in Table 5.2, and they start to build shared understanding of the relevant science content. Thus, we see faculty's interactions become more consistent with scientific sense-making together as the task progresses. Following this pseudo-student experience, faculty identify WL facilitation moves that they found valuable, justify the pedagogical worth in these facilitation moves based on their pseudo-student experiences in this session, and directly point out their initial collaboration troubles. Thus, following all of these interactions, faculty may have become better able to facilitate students' interactions in their own classrooms.

We suggest that these variations in faculty's experiences between the two cases could be linked to variations in the WL's facilitation moves. In the whiteboard task, the WL did not notice moments of low coordination in faculty's group work and did not intervene in pedagogically thoughtful ways. Instead, she briskly corrects faculty as if admonishing them for not following the rules (not drawing the grid quickly enough), even though these rules seemed unclear to faculty. This facilitation move seems analogous to WL-B asking faculty to fill in TPS implementation critiques at times when they are not likely to appear competent in the previous subsection (Table 5.1): in both cases, the WL reprimands faculty at times when they seem unlikely to understand the reasoning behind the corrections. Moreover, the WL did not encourage faculty to reflect on their pseudo-student experiences, leaving all of faculty's attempts to reflect unsupported and unstructured.

In contrast, in the Lecture-Tutorials session, both of the WLs intervened in faculty's pseudo-student interactions in pedagogically rich ways. One WL modeled potentially desirable pedagogy in guiding faculty to articulate and resolve conceptual sticking points, while the other intervened at a point of particularly low coordination and encouraged faculty to work together. In this second case, faculty were able to recognize the purpose of this intervention because it was accurate and relevant for their current collaboration struggles. Again, this seems analogous to a WL facilitation move in the TPS implementation sessions, where WL-C sets faculty up to be successful at analyzing instruction by choosing an accessible example and eliciting their responses. Here, in both cases, the WL scaffolds faculty's learning by pointing out corrections that are likely to appear sensible to faculty. Beyond modeling useful instructor moves, the lead WL also created some protected space for faculty to reflect on their pseudo-student experience: before initiating the science task, he told faculty they would consider strategy implementation, and following the task, he asked faculty to reflect together in small groups and then share out.

As in the previous section, we see evidence that given adequate support, faculty can engage in pedagogical sense-making following experiences of simulated instruction in the workshop. However, we again see only limited evidence of faculty providing robust pedagogical justifications, potentially because of time constraints enforced by the WLs (at least in part). In the next and final episode, we will look for evidence of faculty providing these more elaborated justifications.

5.6 Faculty watching classroom video

We have argued at earlier points that faculty have potentially productive ideas that could help them to make sense of instructional decisions, and have found some limited evidence to support this so far. However, we have yet to identify an episode of extended talk in the NFW where faculty have space to make complex or nuanced arguments about instructional decisions without substantive input from a WL. Here, we turn to a session in which we expect to see extensive faculty talk and pedagogical reasoning based on the high percentage of dialogic engagement captured with our workshop observation tool. (We describe our initial hypotheses about this session based on R-PDOT data in Chapter 4, where it is identified as "Session C.") While there is both extensive small group discussion and large group open discussion in this session, we consider one episode of large group discussion (with about 20 faculty participants) to illustrate how the WL interacts with faculty and facilitates this open discussion. In the discussion, several faculty appear to summarize ideas that came up in their small group conversations, so this likely gives us insight into the ideas that came up in multiple groups in addition to demonstrating what the emerging faculty interactions look like.

Specifically, in this session, faculty analyze a segment of classroom video guided by a transcript, a representation of the task that students in the video are working on, and brief guiding prompts. This activity is taken from a larger library of video resources and associated activities, which the WL describes at the start of this session:

> Periscope are a series of lessons that are centered on video episodes from best-practices classrooms. And they're supposed to help instructors with noticing and interpreting student behavior, and to practice lessons they've learned about teaching in actual teaching situations. ... [T]he goal is to help instructors see authentic teaching like an expert educator does, to develop their professional vision of what it means to be a teacher.

We note that the WL frames these videos as opportunities for instructors to learn to notice and interpret student behaviors in more expert-like ways. It is not clear if faculty will see themselves as "instructors" who are supposed to be learning in the context of this session, since this is framed very generally. Still, although she does not presents these as goals for this NFW session, her statements are closely related to our discussion of ambitious goals for PD, and video lessons like the one faculty will experience here are explicitly designed with the intention of support these goals. We also note that framing these videos as "best-practices classrooms" may encourage faculty to find value in the behaviors of the instructors in the video from the onset—there may be some associated expectation that the instructors in this video are modeling what the designers perceive to be "best-practices."

Following this brief introduction, the WL launches into a specific video lesson by asking faculty to read and discuss the task that students in the video will be working on (using a paper handout they are provided), telling faculty "I'd like you to just discuss with your partner how you think the students might answer that." By doing this, she gives faculty some protected time to make sense of the physics task and consider ideas that students might have related to it. After letting faculty discuss in small groups for a couple minutes, she plays the video, stating "I'm just gonna ask you afterwards to talk with your partner about what you noticed, very open ended question." She plays the 3-minute video in which a physics TA (Levi) interacts with a group of students who are working on a physics Tutorial about how pressure operates in fluids. After watching the video, faculty discuss with their partners for a few minutes, and then the WL begins a follow-up conversation with all participants as follows.

WL: [*talking over FP*] Okay, I'd love to hear what you got. Let's come back. (*FP become quiet.*) What'd you see?
A: He didn't say anything right away, Levi didn't. He waited a solid thirty seconds maybe before he jumped in.
WL: What'd you think of that? A: I think that's probably a good thing to at least understand what the discussion (is), from his perspective.
B: I think the prompt is a little complex in that it's got both depth and size of container, and so from the level of understanding that the students have had since the beginning, they're still working about the definition of h. Maybe they could have simplified.
WL: So that could have been too complex for them to (inaudible).
C: It kinda surprised me, that it seemed like one student clearly knew what was going on and could have helped the others. And she did, but it sorta surprised me that she was being polite about it and not so much saying "Listen you're wrong, this is how you should be thinking about this." She was kinda giving them equal time. And so it was hard, it was probably hard for the other students to realize "Oh she actually know actually knows what's going on and she's maybe just being polite."
WL: That was Alicia, or Cass?
C: I think it was Alicia.
D: Cass is interesting one because she might be the one that understands the most, but she's not saying it. It's difficult to understand whether even she is understanding or not, because she's not participating. And I like the way Levi tried to bring her, like, "You seem to disagree, to what do you disagree?" So he's trying to bring Cass into the discussion.
E: So it seemed like there was definitely some definitional issues that the students were dealing with, and maybe that came from lecture, and we don't know what happened in lecture so we don't know if the professor just did a poor job or if it was intentional, but what I liked was that Levi just put it aside. He was like "Lecture's not, I wasn't in lecture, let's just tackle this problem with our brains and think about it," which I thought was really great, and then you're not kinda hinging everything on what the professor said in class.
F: I think one thing I really appreciated was he came and sat down and didn't say a single word until one of the students said "I have a question," and then like you said he didn't introduce anything he just waited until "I have a question."
G: We noticed that what was also good about the TA was he didn't tell them the answer, but instead guided them to formulate their claim in a complete sentence, so he guided them into formulating their idea even though it was wrong, and refrained from teaching them anything apart from helping them to come up with some kind of claim by themselves.
WL: Was anyone concerned about that, that he was helping them guide the formulation but it wasn't going to be correct necessarily?

A: Maybe a little. I mean I noticed at one point he repeated back some specific language that they used, and that's probably not how I would describe it and I'm not sure it's a great way to describe it so I don't know if, I don't know if maybe he should have tried to steer them in a different direction.

Several aspects of this discussion are noteworthy to us. Overall, we find this to be a rich conversation, in that multiple faculty participants are evaluating instruction and providing pedagogical justifications to support their evaluation. We can hypothesize about what might have supported this pedagogical sense-making within the above interaction by first unpacking how the WL talks and acts. The prompt that initiates this large group discussion is very open ("What'd you see?"), which matches the initial prompt that launched faculty's small group discussions. When faculty A responds with an observation of the TA's behavior, the WL responds with another open, follow-up prompt ("What'd you think of that?") that invites faculty A to evaluate this behavior himself. Following this, we note that the WL says very little for several turns of talk, simply acknowledging and calling on faculty to voice their ideas without interjecting her own ideas, and helping faculty to clarify their thinking for others (e.g., "That was Alicia, or Cass?"). It is only at the very end of this interaction that the WL substantively alters the flow of the conversation by providing a reason that faculty might disagree with one another and inviting them to voice opinions along these lines. In this last case, it seems that the WL gives faculty social permission to disagree both with one another and with the TA's decision in the video, which one faculty member takes up. Prior to this, most faculty praised the TA and did not disagree with their peers within this discussion.

In order to qualify our claim that a rich pedagogical discussion unfolds here, we consider the substance of what faculty are saying. First, we note that after the first interaction, most faculty automatically go beyond noticing to make evaluative statements about the TA's instructional moves, such as, "I like the way Levi...," "I thought [the TA's move] was really great," "one thing I really appreciated...," and "we noticed that what was also good about the TA...." Several faculty (A, F, G) notice and praise the fact that the TA does not say anything when he first sits down with the students, and instead waits until they ask a question. Similarly, faculty notice and praise the fact that the TA does not tell students the answer but rather tries to pull students into the discussion (D) and encourages students to "come up with some kind of claim by themselves" (G) and "tackle this problem with our brains and think about it" (F). We find these statements to be strongly aligned with education research recommendations and principles, and observe that they sharply contrast with conceptions of teaching-as-telling that some researchers and PD leaders might expect these new faculty to hold to the exclusion of all other forms of teaching. (Since it is certainly possible that these faculty do not have exclusively "expert-like" ideas about teaching, our more nuanced and flexible picture of faculty's thinking might better explain what we observe here.) We also notice that while these ideas may represent what faculty had discussed in their small groups, faculty seem to be listening to and leveraging the ideas voiced in this large group discussion at least to some extent, e.g., "like you said..." (F), which suggests that at least some faculty are carefully listening to their peers and possibly learning from them.

Faculty also notice aspects of student behaviors that could support them in thoughtfully facilitating group work. For example, C suggests that students may not recognize each other's expertise because of the way they choose to interact with each other, e.g., he suggests that one student refrains from strongly disagreeing with her peers in an effort to be "polite," and this may have limited their progress on the task. Faculty D notices when another student says very little, "it is difficult to understand even whether she is understanding or not," and identifies a specific instructional move on the part of the TA ("you seem to disagree, to what do you disagree?") that may have ameliorated this collaboration challenge. Faculty use concrete examples to anchor their points, but we see how faculty's ideas could be generalized to noticing and responding to student behavior in similar situations. In other words, while faculty do not explicitly state that these ideas are applicable or relevant in other situations, they might recognize these similarities or could be guided to do so.

5.7 Conclusions

Within the context of the NFW, all of the sessions we have analyzed in this chapter provide rare opportunities for faculty to experience and analyze concrete examples of instruction. While we identified some workshop leader design choices that seemed less fruitful than others, we recognize that facilitating productive faculty conversations within the NFW could be challenging for structural reasons beyond an individual workshop leader's control. For one, we note that workshop leaders face strong time constraints that may be detrimental to their facilitation moves and design choices. Multiple workshop leaders have said this to us in interviews and informal conversations, and we see this evidenced in these transcripts when workshop leaders mention time constraints when introducing tasks to faculty. Moreover, some of the school-like orientations and ideas that faculty bring into these particular sessions, which make it more difficult for them to learn, may be connected to prior experiences in the NFW. Our episode-specific analysis has intrinsic limitations, as we can only carefully model small pieces of faculty's experiences that are embedded in a broader context. However, our field notes suggest that "doing school" could already be cued up based on other workshop leaders' didactic and highly prescriptive descriptions of RBIS. Thus, just as classroom educators may struggle against broader cultural norms when teaching students, workshop leaders may face challenges in encouraging faculty to collaborate and reason about instruction because of faculty's long ago experiences as students, their more recent experiences as instructors, and their very recent experiences in other NFW sessions.

At the same time, not all of these sessions appear to promote pedagogical sense-making to the same extent, and variations in the quality of faculty's interactions across similar sessions provide valuable insight into how these kinds of sessions can be productively structured and facilitated. Through our qualitative analysis, we were able to build design conjectures that describe how session embodiment is linked to mediating processes within each of the three contexts where faculty experience and reflect on concrete examples of instruction in situ.

We first considered situations where faculty act as instructors to simulate a prescribed RBIS while their peers act as pseudo-students, and this pseudo-instructor implementation is critiqued. Based on our data, we developed a pair of design conjectures we summarize as:

– If the basic session structure follows the Prather et al. [2009] model and the workshop leader's facilitation moves focus on

  (1) enforcing predetermined behaviors that seem opaque to faculty,

  (2) highlighting the workshop leader's own authority and expertise, and

  (3) evaluating the correctness of implementation based on alignment with the prescribed script,

  then faculty will engage in "doing school," going through the motions without articulating any underlying pedagogical logic.

– On the other hand, if the basic session structure again follows the Prather et al. [2009] model but the workshop leader's facilitation moves focus on

  (1) responding to and gently navigating faculty's incoming expectations and affective responses,

  (2) developing a sense that "we're in it together" where power differentials are minimized, and

  (3) encouraging norms for exploring alternative instructional moves and providing pedagogical logic behind implementation choices,

  then faculty will engage in pedagogical sense-making and articulate implications of instructional choices for student engagement and learning.

Next, we considered situations where faculty act as pseudo-students while the workshop leader(s) simulate the role of a college science instructor. Here, we developed the parallel conjectures:

– If faculty are presented with a challenging scientific task, and workshop leaders

(1) pay limited attention to the quality of faculty's small group collaboration and their own redirection of faculty's focus during a science task

(2) do not create protected space for faculty to reflect on their experiences,

then faculty may struggle to collaborate smoothly and build scientific meaning together, and will struggle to pursue reflective discussions together.

– Conversely, if faculty are presented with a challenging scientific task, and workshop leaders

(1) model potentially desirable pedagogy and

(2) create opportunities for faculty to reflect on their pseudo-student experiences,

then faculty will be able to repair moments of low coordination and build scientific meaning with their peers, and will identify facilitation moves they might try out in their own classrooms.

Lastly, we looked at one case where faculty watch and analyze classroom video together. While it is more difficult to build strong conjectures from a single case, we developed the initial conjecture that:

– If...

(1) the video is relatively short, focused on a conceptually rich student task with accessible and relevant disciplinary content, and includes rich interactions among students and between students and an educator,

(2) driving questions prompt faculty to reflect on what they see happening,

(3) faculty are given protected time to make sense of the physics and their initial impressions of the episode in small groups, with transcript pro-

vided,

(4) the workshop leader's facilitation moves invite faculty to share and elaborate on their own ideas and respond to each other's ideas,

then faculty will (as above) engage in pedagogical sense-making and articulate implications of instructional choices for student engagement and learning.

This research has some direct implications for workshop design. We see evidence that each of these kinds of workshop activities has the potential to foster faculty's productive engagement around analyzing instruction, and therefore are potentially valuable components for workshop designers to include. If faculty engage in this pedagogical sense-making in workshops, they may be better able to adapt to emergent situations in their own classrooms and have fruitful conversations with other educators when they return to their home institutions. While we have not explored the different affordances and limitations of these different ways to simulate instruction in workshops relative to one another nor how variations in the task itself might influence faculty's engagement, we suggest that all three of the tasks we observed might complement each other within an extended workshop.

However, we have also shown that not all enactments of these workshop tasks will support faculty's engagement in pedagogical sense-making: the workshop leader's implementation matters. Faculty may not collaborate smoothly and may easily take up school-like norms, which can be cued up or reinforced by a workshop leader. However, when struggling to collaborate with their peers, faculty may respond well to a workshop leader's sensible and relevant intervention, which suggests that workshop leaders could benefit from looking for and responding to markers of low collaboration in faculty's interactions (Table 5.2). We also find evidence that faculty may spontaneously notice and evaluate salient aspects of the instructional examples they experience in workshops and provide pedagogical justifications to support possible instructional choices, even without direct prompts from workshop leaders. Workshop leaders could draw out these ideas when they come up and make dedicated space for these reflective conversations to occur. In doing so, our data suggest that workshop leaders may enable faculty to voice and build up pedagogical ideas that are highly aligned with education research principles, thus creating rich opportunities for faculty to learn together.

More broadly, some workshop leader framing and facilitation moves could cue up and sustain faculty interactions that are consistent with "doing school," while others seem more likely to support pedagogical sense-making. While some of the facilitation moves we summarize in Table 5.1 are narrowly applicable to the context of faculty simulating RBIS, similar workshop leader facilitation moves could be similarly consequential in all workshop contexts, as we demonstrate to a more limited extent in the other episodes. In particular, we note that a workshop leader's exclusive focus on adoption of prescribed RBIS may risk cuing up school-like norms in faculty, while allowing faculty to consider the possible affordances and drawbacks of RBIS adaptations may naturally lead to pedagogical sense-making.

Thus, we recommend that workshop facilitators:

– Create extended opportunities for faculty to both experience and reflect on simulated RBIS together;

– Model pedagogical sense-making by encouraging faculty to reason about when and why adaptations to prescribed RBIS might be appropriate;

– Work to minimize power differentials between themselves and faculty participants;

– Look for markers of faculty beginning to engage in pedagogical sense-making, and if these markers emerge, invite faculty to elaborate on their reasoning and contest each other's ideas, or highlight the potential productivity in these ideas;

– Look for markers of faculty "doing school" with respect to learning about instruction (e.g., focusing on learning prescribed rules), and if these markers are prevalent, press faculty to consider the underlying pedagogical logic behind these instructional choices;

– Look for markers of faculty "doing school" when acting as pseudo-students, and if these markers emerge, intervene to encourage more productive collaboration and disciplinary sense-making.

Although we consider these initial results to be potentially useful to workshop leaders, from a research perspective, there is still much left to explore. The nature of faculty's interactions around instruction and the supports necessary to foster this kind of learning present highly under-explored areas of research. The conjectures we developed are intended to be iterated on, as they open up questions about which of these design features are the most consequential for faculty's engagement and whether key design features are still unspecified. Similarly, we have not captured a full range of variations that could either support or constrain faculty's learning about instruction within these contexts, nor have we captured all contexts in which faculty could reason about instruction. We also have not empirically investigated how these mediating processes might lead to various outcomes or built specific theoretical conjectures through this type of investigation. Therefore, from a research perspective, we aim to provide initial models for PD design that future researchers can expand on, complement, and refine, in order to build more robust models of how to support faculty's learning in PD contexts.

Chapter 6: The transformation of an introductory astronomy course sequence for majors: a local, team-based approach to instructional change

6.1 Introduction

In the final body chapter of this thesis, we will turn to a local course transformation effort in the University of Maryland Astronomy department, in which I played a central advisory role. For readers who are beginning at this point, I will note that we have been discussing research-based instructional strategies (RBIS) throughout this thesis, that I motivated my own interest in RBIS in Chapter 1, and that attention to RBIS aligns with national priorities and recommendations. In particular, there is broad recognition at the national level that widespread incorporation of these instructional strategies into undergraduate science, technology, engineering and mathematics (STEM) courses has the potential to increase participation in these fields [President's Council of Advisors on Science and Technology, 2012, Singer et al., 2012, National Science and Technology Council Committee on STEM Education, 2013]. Similarly, national attention has been paid to increasing equity and inclusion within STEM, as it is clearly evident that demographics in

STEM fields are not representative of the U.S. population [Hill et al., 2009, Merner, 2015, Mulvey and Nicholson, 2015]. Research studies have demonstrated that while more equitable outcomes are not guaranteed, RBIS have the potential to improve student learning, attitudes, and odds of passing STEM courses for all students, not just for those students who are over-represented relative to the U.S. population [Rudolph et al., 2010, Freeman et al., 2014, Brewe et al., 2010], and have been shown to eliminate the gender/ethnicity gap in student pass rates in the most promising cases [Brewe et al., 2010].

Among the national astronomy community, there is significant momentum towards making our field more inclusive (e.g., [Urry, 2003]), which could easily translate to interest in implementing active learning strategies in the classroom. Significant resources exist for implementing active learning at the non-majors "Astro 101" level: astronomy education researchers have developed and disseminated RBIS such as Lecture-Tutorials, interactive simulations, and multiple-choice questions that can promote student discussion and improve students' conceptual understanding of astronomy during class [Prather et al., 2004, Prather and Brissenden, 2008, Hudgins et al., 2006, Prather et al., 2009]. Similarly, numerous instruments have been developed to assess students' conceptual understanding and demonstrate the utility of these RBIS in non-majors' astronomy classes [Hufnagel, 2001, Lindell and Olsen, 2002, Bardar et al., 2005, Keller, 2006, Bailey et al., 2011, Williamson and Willoughby, 2012, Slater, 2015].

Despite this wealth of resources for Astro 101, few curricular innovations have been developed specifically for courses for astronomy majors (Christian and Belloni

[2003] is an exception, and Sanders et al. [2012] provides some starting points), and little to no formal research has been conducted on active learning in astronomy majors' courses [Deustua et al., 2009]. If a goal of our community is to create equitable pathways for future astronomers, this points to a significant gap in our understanding.

The lack of research-based curricular materials specific to teaching astronomy majors does not imply that no relevant research-based resources exist, but it does suggest that significant work is likely needed to adapt these resources to fit within this context. Some instructional building blocks could be adapted from research-based instructional strategies curricular materials in physics, but much of the typical content in majors' courses is unique to astronomy, and thus will require the use of some astronomy-specific instructional materials and assessment instruments. Similarly, resources for teaching astronomy majors can be drawn from existing Astro 101 materials, but there are potential limitations to doing this and modifications are likely needed. Most pressingly, while the topics in introductory astronomy courses for majors likely overlap with topics included in non-majors' classes, the level of math and physics preparation might be appropriately assumed to be higher, and we lack a strong basis for understanding how astronomy majors will make sense of Astro 101-level materials alongside more advanced topics and mathematical problem-solving.

Trends in instructional goals for majors-level astronomy courses may also differ from trends in the goals for Astro 101, which may further influence what curricular materials have been developed extensively. There is significant variation in which content is taught in Astro 101 [Slater et al., 2001], and instructors' overarching course goals for non-majors are often not content-specific, such as helping students to become more informed citizens [Wallace et al., 2013] or improving their mathematical reasoning abilities Follette et al. [2015]. Therefore, in Astro 101, instructors may consider the choice of content less important than the amount of excitement about science or general mathematical literacy they are able to foster, and may not perceive a need to align their goals with other instructors' goals for (nominally) the same course. In contrast, courses for astronomy majors likely aim to teach skills and knowledge that would be particularly useful to students who are considering the pursuit of astronomy-focused careers, and this will likely shape and constrain instructors' choices differently. When considering courses for astronomy majors, it might be easier for multiple instructors (e.g., within a department) to identify and build consensus around what constitutes foundational content and skills, and to focus on developing students' abilities in these areas across multiple courses. Because astronomy curriculum developers have often worked to develop materials that span many topics in order to meet the diverse goals of Astro 101 faculty, extensions of these curricular materials to explore some foundational topics in more depth are likely needed.

Moreover, while it is arguably important for any science class to foster community among students, this is particularly important and relevant when teaching astronomy majors, whose social connections with their peers could either support or limit their future participation in the field. Students who find supportive communities among their peers may be more likely to persist as astronomy majors. Thus, particular attention to this potential benefit of some research-based instructional choices is warranted when designing learning environments for astronomy majors.

In the first half of this chapter, we outline the model we developed for a two-semester introductory course sequence for astronomy majors at the University of Maryland (ASTR120/121) and the resulting student outcomes. This course sequence encompasses many topics that are typical of an Astro 101 curriculum along with more advanced physics and mathematics, as well as a laboratory component where students are expected to write basic computer code to simulate and analyze astronomical data. In Section 6.2, we describe the ways in which we have restructured the lecture portion of the course and dramatically re-envisioned the laboratory curriculum (which occurs in the second semester, ASTR121). In Section 6.3, we consider the student outcomes we have been able to document, including the results of our use and/or adaptation of existing research-based assessment instruments. While this course transformation was not undertaken as a formal research project and some instructional materials could likely be improved by closely examining students' responses, we aim to give other astronomy instructors new ideas about how to improve instruction for astronomy majors and increase awareness of the initial materials we have developed.

In keeping with the theme of this thesis, we use the second half of this chapter, Section 6.4, to describe potentially salient features of the change process that led to this instructional model. All of our course transformation efforts were enacted by a course transformation team with a broad range of pedagogical and disciplinary expertise, and were enabled by departmental and institution-level supports. We are interested in team-based instructional change as a mechanism for promoting and sustaining instructional improvements, and aim to contribute to the growing research base on how team-based change works by providing this retrospective change account. In a practical sense, by describing our change process, we aim to help other instructors (within astronomy or otherwise) to consider potential supports within their local context that might help them to successfully pursue instructional change in similar ways. Conversely, from an administrative perspective, this description of our change process also suggests how various institutional supports might enable instructional improvements to thrive.

We conclude by summarizing the key findings in this chapter and considering possible directions for future astronomy education researchers in Section 6.5.

6.2 The course structure

The ASTR120 and ASTR121 courses at the University of Maryland are a two-semester introductory sequence required for the astronomy major. Although primarily slated as an introductory sequence for freshmen intending to major in astronomy, ASTR120/121 has also historically been taken by both upper-level physics students intending to add an astronomy double-major, and freshmen declared as non-science majors or with no declared major. The first half of the first semester (ASTR120) introduces fundamental skills in astronomy, which includes: understanding how the apparent motion, brightness, and size of nearby objects in the night sky relate to the Earth's motion and their true sizes and distances; predicting how objects will move under the influence of gravity; and detecting and interpreting the basic properties of light across the electromagnetic spectrum. This is followed by the study of specific objects in planetary science (planets, moons, comets, etc.), including extrasolar planets and the Sun. The focus of the second semester (ASTR121) lies beyond the solar system and focuses on the study of stars, galaxies, and the universe in general. There is a lab component in ASTR121 intended to introduce students to the techniques and tools of professional astronomers, including computer programming and scientific writing.

6.2.1 Previous course structure

Prior to our redesign, ASTR120/121 consisted largely of traditional lecture with interactivity limited to the instructor posing and fielding questions of the entire class in an unstructured way. The textbook was considered optional and was used mostly as a source of lecture notes for the instructor (PowerPoint slides) and some homework questions. The weekly discussion sections consisted of reviewing topics from lecture and assigning worksheets written by the Teaching Assistant (TA), based on limited consultations with the instructor. There were no formal learning goals for these courses, beyond the notion that covering a broad range of topics was considered an essential introduction for the astronomy major, and that the course sequence would require higher-level problem-solving and mathematical skills compared to a single-semester, non-majors introductory astronomy course. Since ASTR120/121 instructors typically want to include material not captured in the introductory astronomy textbook (likely written with Astro 101 students in mind), lecture topics were supplemented by more in-depth notes in some areas as well as mathematical derivations carried out for the class on the front whiteboard, using resources passed down from previous instructors of the course that predated widespread use of PowerPoint.

Although student evaluations for this prior version of the course showed high satisfaction, there are many reasons that we might be dissatisfied with past outcomes. For one, as we mentioned in Chapter 1, exclusive use of traditional lecture puts underrepresented students at a strong disadvantage. Only a few students, typically those in privileged groups within society, tend to be comfortable speaking up in front of the entire class to ask or answer questions, and the subtle ways instructors respond to these students often reinforces these norms [Johnson, 2007]. Because of this, traditional lecture does little to help most students actively refine and build on their existing ideas during class [Bransford et al., 2000, Redish, 1994], and we have indeed noticed disparities in student participation in ASTR120/121. There also appear to be more longitudinal challenges for student success in our department. Anecdotally, students progressing to later courses in the astronomy major have trouble remembering or applying foundational astronomy concepts. Many students also seem to struggle to complete future coursework that requires programming, as reported by both former students and instructors of later courses. Students have explicitly noted significant drawbacks to the lab component of ASTR121: students who lacked preparation in computer programming found the exercises to be difficult and discouraging, while students who were already adept at programming felt bored. We also suspected, based on our past experiences and others' formal research on students' participation in computer science [Margolis and Fisher, 2003], that White male students are more likely to have the prior program experience that offers them an advantage in understanding and quickly completing the lab (leading to boredom instead of feelings of falling behind).

The challenges that students face are not particularly surprising, given that many elements of the course (both the nature of assignments and the use of in-class time) were structured in ways that could reinforce rote memorization of facts and formulaic calculations over critical thinking. The ASTR121 lab left little room for students to develop their programming and problem-solving abilities by exploring different approaches, and little structured time to reflect on past work. Moreover, among the students taking ASTR120/121, there are many with either extensive or minimal mathematics or physics training, and a sizable fraction with incoming GPA below 3.0. When it comes to the ASTR121 lab, we know in general that women have fewer opportunities to learn programming skills before college [Margolis and Fisher, 2003], and we have no reason to believe that students at our institution are an exception. It seems likely that previous versions of ASTR120/121 have done little to reduce disparities in students' preparation before they proceed to later courses, and thus have left many underprepared to meet the expectations of future instructors and advisors. While graduate students had previously been tasked with writing and revising the lab, they did so with little to no support from the lead instructor or other consultants, and had not met with satisfactory improvements in student outcomes. All of these shortcomings and challenges motivated our course transformation effort.

6.2.2 Current course structure

Our redesign proposal for ASTR120/121 was selected for funding in 2015 by the University of Maryland Teaching and Learning Transformation Center (TLTC) as part of their Elevate Fellows program.[1] We established course-level learning goals to provide structure to the redesign process and to guide students' focus during the course (Table 6.1) [Chasteen et al., 2011, Wieman et al., 2010]. Similarly, we articulated learning goals for every individual lecture, as well as each lab in ASTR121. There now are 2–3 learning goals for each typical lecture, provided to the students along with their reading assignments prior to class, and these form the basis of in-class activities, homework questions, discussion topics and worksheets, and exams.

Table 6.1: Course-level learning goals for ASTR120/121.

- Develop an appreciation for our place in the universe.
- Convey the current state of knowledge regarding stars, galaxies, and the universe in general to a non-specialist.
- Solve complex problems requiring application of multiple astrophysical concepts.
- Collaborate with others to develop shared knowledge.
- Write scientifically and communicate your results effectively.
- Critically evaluate your own and your peers' work.
- Interpret error, accuracy, and precision of astrophysical measurements both in the existing literature and for your own data.
- Use MATLAB to analyze and visualize astrophysical data.(ASTR121 only)

With a content focus established by the class-specific learning goals, activities in the redesigned course follow a consistent pattern. Table 6.2 shows examples the instructional elements in a particular ASTR121 lecture. On a typical day, the

[1] A description of the Elevate program can be found here: http://tltc.umd.edu/elevate-program-description.

instructor has assigned students reading prior to each lecture, supported by online low-stakes reading quizzes (sometimes in-class quizzes), so they might arrive to lecture familiar with the basic material and at least minimally prepared for group activities following this initial exposure. In class, the instructor first reviews the learning goals for that day. The instructor often introduces "muddiest points", i.e., students' questions or reflections about content that confused them from an online exercise (quoted anonymously), to guide class discussions at the start of class or as discussions unfold (as in "Just-in-Time Teaching" or JiTT, [Novak et al., 1999]). If there is content that is not included in the textbook but required for a learning goal, a short lecture will sometimes precede more interactive discussions and student work. Occasionally this supplemental content involves a mathematical derivation done by the instructor at the whiteboard (though we now sometimes provide these mathematical derivations in the form of videos, as discussed below).

As class continues, students regularly vote on conceptual, multiple-choice Think-Pair-Share (TPS, also known as Peer Instruction) questions with colored cards to probe where the instructor should spend more time and create opportunities for student discussion [Mazur, 1997, Prather et al., 2009]. The questions used come from a variety of sources, including the University of Nebraska-Lincoln ClassAction website (www.astro.unl.edu), the Lecture-Tutorials for Introductory Astronomy instructor's guide [Prather et al., 2004], and PowerPoint slides and online quizzes (MasteringAstronomy) provided by the textbook authors [Bennett et al., 2016], while some are of our own invention guided by the learning goals. Although most of these sources provide questions that were developed for non-majors, we found that the subset of questions that were the most conceptually challenging for non-majors also typically appeared to be challenging for our majors.

When the instructor implements TPS in ASTR120/121, if many students do not initially vote for the canonically correct answer to a question, the instructor asks them to break into groups to discuss the question among themselves, with the instructor and Teaching Assistant (TA) or Learning Assistant (LA) circulating to facilitate, then asks for a re-vote. If a majority of students vote for the expected answer after peer discussion, a student will typically be asked to explain their reasoning. If there is still significant disagreement, the instructor will sometimes elaborate further using additional PowerPoint slides or writing on the whiteboard while lecturing. These PowerPoint slides would often not be displayed if students came to consensus around the canonical answer without guidance. We note that this aligns with some, though not all, of the steps prescribed by Mazur [1997] and Prather et al. [2009]. While two members of the team were aware of all of the components of the TPS script as described by Prather et al. [2009], there was no particular focus on communicating this to the instructor initially or trying to encourage fidelity of implementation. Therefore we note that other instructors could take different approaches to enacting the "same" instructional strategies in this course sequence, which could have different affordances and/or drawbacks [Turpen and Finkelstein, 2009, 2010, Dancy et al., 2016].

At least once during most lectures the instructor assigns a group whiteboard activity. This instructional strategy is broadly inspired by the use of whiteboards in the Paradigms in Physics project at Oregon State University [Manogue et al., 2001], as well as the introductory physics for life science majors course sequence here at the University of Maryland [Redish et al., 2014]. Students break into groups of 3–4 students to answer a question that usually involves either a calculation or a sketch, and the instructor and a TA facilitate. The calculations are similar to what is assigned in homework or given on exams, and related directly to a learning goal. The sketches or other conceptual questions are also based on the learning goals and could be assessed in a similar way in homework or on exams. Each group has a mini whiteboard and dry-erase markers (a different color for each group member). At the end of the activity, which may take 15 minutes or longer depending on the complexity of the problem, the lead instructor asks the students to take a photo of their whiteboards to upload to our course management system for credit. This may be followed by discussion, particularly if there are any questions stemming from the activity.

The increased interactivity of "lecture" enabled by the course redesign is corroborated by classroom data that was collected using the Classroom Observation Protocol for Undergraduate STEM (COPUS) [Smith et al., 2013] for 3 lectures across Spring 2015 and Fall 2015. We show the most recent COPUS observation in Figure 6.1. In this figure, we see that student group work, while still less prevalent than lecture, takes up a sizeable fraction of the class time, and that both TPS questions (clicker questions) and whiteboard activities (worksheet group/other group) occurred on this day. We can also see that in addition to lecturing and answering questions, the instructor (Derek) both posed TPS questions and questioned the whole class directly for a substantial fraction of the class time, and that he walked around the room to interact with students as they worked with their peers. Based on the evolution of Derek's instruction, we suspect that most COPUS observations in ASTR121 would show similar or higher levels of interactivity.

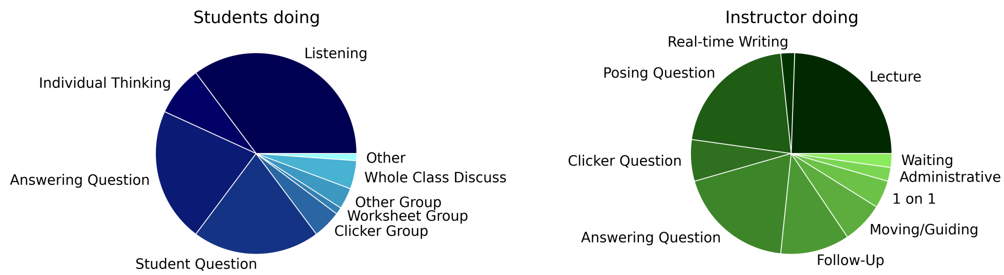

Figure 6.1: Pie charts showing student and instructor behaviors in ASTR120 during the middle of the Fall 2015 semester, as documented with the Classroom Observation Protocol for Undergraduate STEM (COPUS) [Smith et al., 2013]. I collected this COPUS data using the online GORP interface from UC Davis (www.gorp.ucdavis.edu) and data was collected continuously (not binned into 2-minute intervals, as in the original, paper version of the protocol). A description of each COPUS code can be found at www.cwsei.ubc.ca/resources/files/COPUS_protocol.pdf.

Table 6.2: Example instructional elements from the class "The Cosmic Distance Ladder" (ASTR121, class 15). For simplicity, we show only one of the two learning goals for this class.

Instructional element	*Example(s)*
Learning goal	By the end of this class, you should be able to explain conceptually how astronomers use different techniques that we have seen throughout the semester to estimate distances to objects in the universe, from the nearest to the furthest scales, and which techniques are most suitable for which distances.
Muddiest points (student questions based on reading)	-How do the rungs in the distance ladder fit together? How can parallax affect [sic] distant standards that only rely on luminosity? -How standard are standard candles? Can a distant B star look like a closer G star? -How does main-sequence fitting work? -Where are Cepheid variables found? Do other stars have period-luminosity relationships? -Why don't we see white-dwarf supernovae in nearby galaxies?

Continued on next page...

Instructional element	*Example(s)*
TPS question	The Hipparcos satellite measured the parallaxes of over 100,000 stars in our galaxy. Suppose there was a systematic error in the measurements such that each measured parallax angle was too large. How many of the following statements might then be true? - Nearby stars would be closer than previously thought. - Distant stars would be farther than previously thought. - Nearby galaxies would be closer than previously thought. - Nearby galaxies would be farther than previously thought. - Distant galaxies would be farther than previously thought. Vote: A=0, B=1, C=2, D=3, E=4.
TPS question	It is possible that the peak brightness of Type Ia supernovae may depend on metallicity (the proportion of heavy elements in the white dwarf). Suppose a higher metallicity leads to a lower peak brightness, and that the actual metallicity is difficult if not impossible to measure. The consequence of this would be: A. Type Ia supernovae would all be more distant than previously measured. B. Type Ia supernovae would all be closer than previously measured. C. There would be increased uncertainty in the distance of all Type Ia supernovae. D. There would be increased uncertainty in the distance of only some Type Ia supernovae.
Whiteboard activity	Type Ia (white-dwarf) supernovae have a peak luminosity of $\approx 4 \times 10^{36}$ W. What is the approximate distance to a white-dwarf supernova with a lightcurve that is observed to peak at apparent magnitude +10?
Lecture elements not in textbook	Tully-Fisher and Faber-Jackson relations for estimating distances to galaxies.

We have taken several other steps to support student learning in lectures and throughout the course, paying particular attention to supports that we think might help students who are typically minoritized and underrepresented within our field, and/or have less incoming preparation. We think that modeling mathematical derivations and problem solving could be a mechanism for preparing students to succeed in future astronomy courses where they will be expected to demonstrate these particular skills themselves [Schoenfeld, 1987]. While many of these exact derivations and problem-solving strategies will be new to all students, some students may have had less practice with this kind of reasoning than others. In an attempt to help students to gain comfort and familiarity with this process, we made video recordings of the more difficult derivations and exercises for students to watch outside of class. In this way, we aim to create opportunities for students to take their time to understand the mechanics of the procedure, replaying it as necessary. We often included problems similar to those demonstrated via video in homework assignments and on exams.

We have also implemented a group retest policy in which students can redo their midterm exams outside of class for partial credit. This approach is largely intended to alleviate stereotype threat, a psychological phenomenon where a portion of a person's working memory is taken up with thoughts of trying to disconfirm negative stereotypes against their social group, which can worsen underrepresented students' exam performance such that their overall score does not accurately reflect their ability to learn or master the course content [Beilock et al., 2007, Steele, 1997, 2010]. The instructor encourages students to work together on these take-home exams, which is consistent with our goal of promoting collaboration throughout the course. We note that students who score lowest on exams initially seem to improve their scores by the greatest amount on through the retest, and that several of our underrepresented minority students (who are more likely to be susceptible to stereotype threat than students who are over-represented in the class) have opted to complete the retest and have improved their scores. While we have insufficient data to examine these trends more robustly, we are encouraged that this policy does appear to be helping some of the students we most want to help. Finally, we have student volunteers available for tutoring throughout each week during the course semester in an effort to provide a friendly peer-level resource for students seeking assistance and a study group outside of traditional office hours.

Our redesign applied to other structured elements of the course as well. The weekly discussion is now tailored to the lecture learning goals, and our team meets weekly to strategize how best to help students achieve these goals. We make extensive use of activities from the Lecture-Tutorials book, which have been developed and refined based on research of student thinking in introductory astronomy for non-majors and have been shown to improve student outcomes in those settings [Prather et al., 2004]. In order to match the higher difficulty level of the course relative to introductory non-majors courses, we only assign the more challenging Lecture-Tutorials. We have found through trial-and-error that while many of the Lecture-Tutorials have worked well in this setting and stimulated rich discussions among students, a few that we tried initially seemed too easy to capture the attention of our students, and we have since stopped using them. Our course transformation team has also generated new, sometimes mathematically driven activities to augment the Lecture-Tutorials or address course topics for which no Lecture-Tutorials exist at the appropriate level. The TA/LA runs the discussion and facilitates student group work on these activities. The discussion is also used for review sessions prior to exams (informed by collecting "muddiest point" responses from student groups) and for collecting research-based assessment data (see Section 6.3).

The lab component of ASTR121 was dramatically redesigned to provide a more engaging and equitable experience for the students. We cut the topics in half, providing 2 weeks per lab, including a more focused training sequence near the beginning for the programming language that we use throughout the semester, MATLAB. The labs are run in two 2-hour sections each week and are led by a TA or, when possible, an LA. Each lab has specific content learning goals and MATLAB learning goals (see Table 6.3), where the MATLAB skills students are expected to learn are cumulative (i.e., they are expected to use early skills in later labs). Compared to the previous version of this lab, there are significantly fewer explicit instructions and more guiding questions, giving students freedom to explore and encouraging them to think about their approach. Pre-lab activities include reading "Astrobites" articles that present research results relevant to each lab.[2] Students work in pairs to complete each lab, and the instructor encourages pairs to compare notes with other pairs, fostering collaboration across the class. During the second week of each lab, pairs trade their draft lab reports for critiquing, giving students practice with the peer-review process. Since many students have little-to-no familiarity with writing professional lab reports with formal sections (abstract, introduction, methodology, results, conclusions), we use our own version of "faded scaffolding" [McNeill et al., 2006, Slater et al., 2010] in which students are guided to appropriately fill in the blanks for a mostly completed report for the first lab project, and gradually write a greater proportion of the reports in later labs, until they get to the final lab report, which they write entirely themselves. The reports are graded using a detailed rubric that is provided to students at the beginning of

[2] Astrobites.org is an online digest with articles written by graduate students, a majority of which summarize current astronomy research studies at a level appropriate for undergraduate physical science majors. Several instructors have used this resource in the classroom previously and other suggestions beyond these have been proposed—see [Sanders et al., 2012].

the course and with the written feedback for each lab. The entire lab curriculum for the 2015 ASTR121 lab is freely available online at ter.ps/astr121lab, and an instructor's guide is available by request.

Table 6.3: MATLAB learning goals for the ASTR121 lab.

Lab #	*Learning goals*
1a (intro to MATLAB)	- Use MATLAB to generate, visualize, and analyze data - Write self-sufficient, organized, and commented functions and scripts - Teach yourself additional MATLAB skills
1b (intro to MATLAB)	- Generate data using a mathematical function and random numbers - Read in data from a file - Explore possible underlying mathematical models representing numerical data
2	- Create a function in a program file - Create and use arrays - Plot data with error bars
3	- Plot functions, overlay multiple plots, and format them - Read in data and define variables
4	- Read in and use data from tab-delimited files - Create plots to clearly display several data sets at once
5	- Plot data with x and y error bars - Fit experimental data and determine information from a fit
6	- Open compressed (.gz) files - Read in data from files in FITS (Flexible Image Transport System) format - Fit experimental data with a custom equation

6.3 Student outcomes

To determine the effectiveness of the ASTR120/121 course redesign, we assessed student outcomes in several ways, both qualitative and quantitative. The quantitative measures indicate a positive overall shift in student learning and attitudes. Qualitative data corroborates these findings: the students are receptive to the new teaching style, see value in group work, and find it useful to monitor and reflect on their own learning and think about how to make adjustments. Details of our course assessments are presented below.

6.3.1 LSCI data

We used the Light and Spectroscopy Concept Inventory (LSCI; [Bardar et al., 2005]) to assess students' understanding of these fundamental astronomy topics. In addition to the topical relevance for these courses, we chose the LSCI out of the existing research-based assessments in astronomy because it has been broadly used to document student learning within the Astro 101 community, which will allow us to contextualize our results within prior work. In the assessment, students are asked to answer 26 conceptually challenging multiple-choice questions related to light and spectroscopy, along with 3 demographic questions. Students were informed that their scores on these assessments would have no bearing on their course grades.

We measured normalized learning gains and effect sizes between the beginning and end (pre and post) of ASTR120 in Fall 2014 and Fall 2015. The normalized learning gain is defined as the difference in the students' mean pre- and post-course scores (the percentage of total items answered correctly) normalized by the maximum possible improvement, i.e., the difference between a perfect score (100%) and the pre-score average [Hake, 1998, Prather et al., 2009]. The effect size is similar except instead of normalizing by the maximum possible improvement, the pre and post score difference is divided by the standard deviation of the entire set of pre and post scores [Cohen, 1988]. The effect size therefore adjusts for class size, but has the additional requirement that only students who took both exams can be counted.

For ASTR120 in Fall 2014, the normalized gain was 0.29 and the effect size was 0.84. For Fall 2015, these values were 0.30 and 0.83 respectively. Compared to a national study of student learning in Astro 101 [Prather and Brissenden, 2008], these gains are consistent with courses reported to be more than 25% interactive (normalized gain of 0.29 or better). The effect sizes are considered large [Cohen, 1988]. By these measures there was little difference in performance on the LSCI between 2014 and 2015, despite the fact that many of the redesign components were implemented in ASTR120 by the end of Fall 2015 (and not during Fall 2014). This could be due to a number of factors, but it is encouraging to note the scores did not *decrease* as a result of the instructional changes.

With regard to Prather et al. [2009]'s national study mentioned above, we note that the similar normalized gains we observe do not imply that astronomy majors left our course with similar levels of conceptual knowledge regarding these topics as non-majors, as the average pre-class LSCI scores in ASTR120 Fall 2014 and 2015—41% and 46%, respectively, as discussed in the text below—were significantly higher than the average pre-class scores measured in any of the Astro 101 courses in Prather and Brissenden [2008]'s study, which were tightly clustered at around 25%. Thus, even though our students had less room to improve because they entered the class with significantly more knowledge about light and spectroscopy, we still observed comparable normalized learning gains relative to highly interactive non-majors' classes, implying that our students also left ASTR120 (and ASTR121) with significantly more knowledge about light and spectroscopy than Astro 101 students following instruction.

We also measured the normalized gain and effect size between pre-ASTR120 Fall 2015 and post-ASTR121 Spring 2016 in order to see if students' conceptual understanding continued to improve in the second semester. These results are represented in Figure 6.2, which we discuss further below. All students who took ASTR121 in Spring 2016 took ASTR120 in Fall 2015 and are represented in this dataset. The normalized gain and effect size were 0.47 and 1.32, respectively, indicating increased mastery of concepts measured by the LSCI during ASTR121. Although we think students could have gained the ability to correctly answer all of the LSCI items during ASTR120 (i.e., there were no relevant concepts left entirely unaddressed during ASTR120), it is not surprising to us that students would develop increased proficiency with these ideas and skills during ASTR121: the LSCI includes many items relevant to light emission from stars, and stellar spectroscopy and evolution were major topics in ASTR121. For completeness, we note that the normalized gain and effect size post-ASTR120 to post-ASTR121 were 0.25 and 0.72, respectively, in 2014/15.

In order to lend additional insight into how students were responding to the LSCI items and how their responses were shifting over time, in Figure 6.2 we show the distribution of student scores at the start and end of ASTR120 in Fall 2015, and at the end of ASTR121 in Spring 2016. Diamonds indicate the mean values for each course. Figure 6.3 shows the percentage of students answering each LSCI item correctly over the same time frame, grouped by item number and sorted in ascending order of percent of students answering correctly in the ASTR121 assessment. The distribution of pre-course responses in Fall 2015 indicates a range in prior knowledge

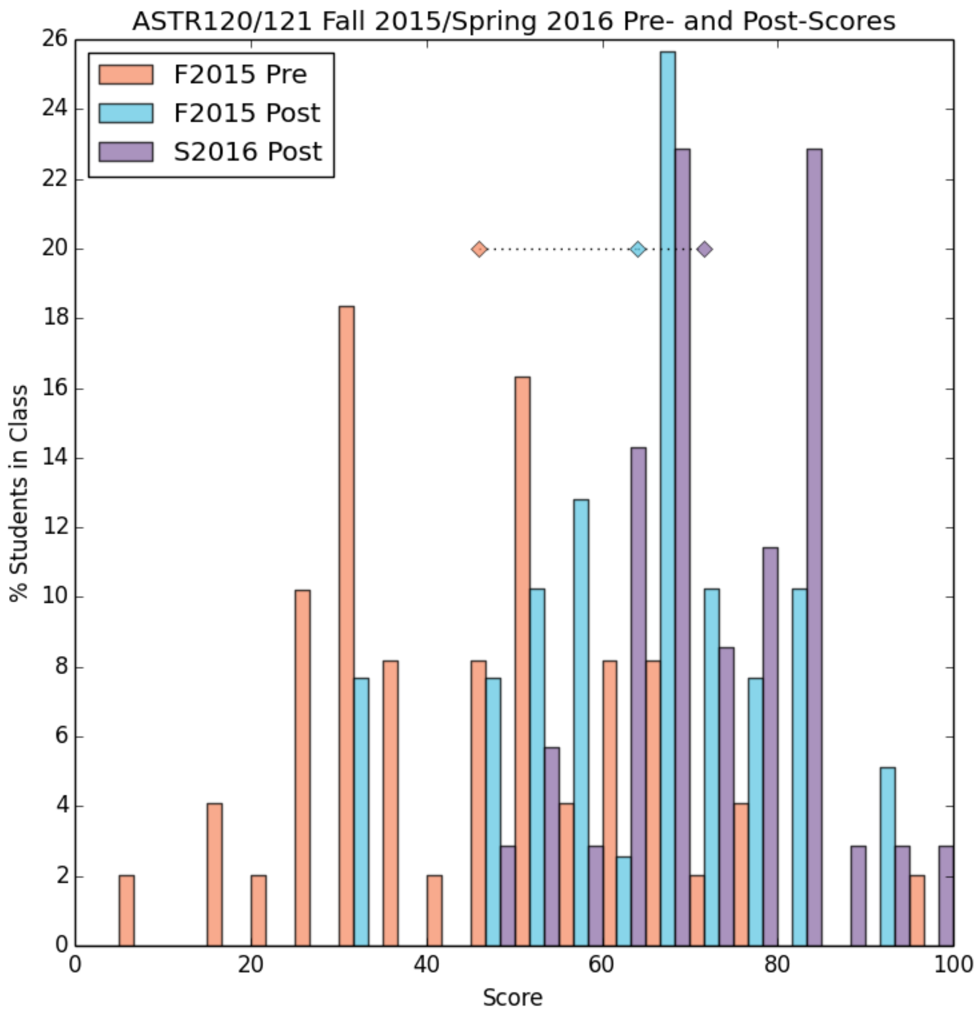

Figure 6.2: Distribution of LSCI scores at the start (orange) and end (blue) of ASTR120 in Fall 2015 and at the end of ASTR121 in Spring 2016 (purple). The diamonds indicate the average score for each course.

for the incoming students (about half of the items had 50% or higher correct answers while the other half had 50% or lower). Generally, scores improved throughout the year, as would be expected, with average scores increasing from 46% at the start of the Fall 2015 semester, to 64% at the end of the Fall 2015 semester, to 72% at the end of the Spring 2016 semester. However we can see that some items are still universally challenging to students and align with what we might expect based on prior research in Astro 101 contexts. In particular, several of the trends we observe in our students' responses are consistent with what other researchers found in an item-by-item analysis of the LSCI. Specifically, [Schlingman et al., 2012]'s analysis suggests that we should expect to observe low post-instruction scores for items 3, 21, and 25, which do indeed appear in the left-hand third of the ranked items in Figure 6.3, and high post-instruction scores for item 14, which is the item that the most students answer correctly following ASTR121 (appearing farthest to the right in Figure 6.3). In this prior analysis, Schlingman et al. [2012] suggest that only item 25 seemed truly problematic for assessing students' conceptual understanding, as students seemed to primarily struggle with terminology when responding to this item. Together, these results suggest that the LSCI is a useful instrument for assessing the knowledge of astronomy majors through the end of their first year of coursework.

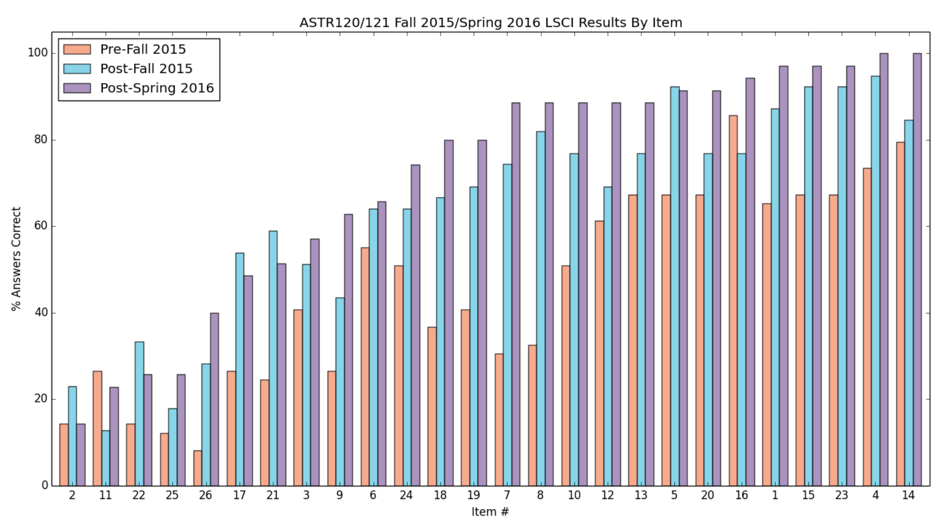

Figure 6.3: LSCI results by item shown as a percentage of items answered correctly, for data collected at the start (orange) and end (blue) of ASTR120 in Fall 2015 and at the end of ASTR121 in Spring 2016 (purple). Results are ordered by percent correct on the ASTR121 assessment.

The average pre-class score was lower for Fall 2014 (41% compared to 48%) but rose a comparable amount (to 58% compared to 64%). This is consistent with the nearly identical normalized gain and effect size measures reported earlier.

While there were not a sufficient number of students enrolled in the class to allow us to draw robust statistical conclusions by disaggregating the data by gender, race/ethnicity, or other demographic variables, we are very interested in whether or not traditionally underrepresented and/or less prepared students are benefiting from the course redesign. Since we had collected information about students' gender, we did check to see if there were any troubling indicators of gender *in*equity within the LSCI data or students' course grades (such as only male students achieving high learning gains and high course grades), and did not find anything obviously problematic. More data would be needed to pursue these issues in depth.

6.3.2 Stop-go-change surveys

6.3.2.1 ASTR120 Fall 2015

As a qualitative measure of student outcomes, in ASTR120 Fall 2015 we distributed mid-semester "stop-go-change" surveys in which students were asked to identify one element of the course that they do not like ("stop"), one that they like ("go"), and one trait about their own learning they would like to modify ("change"). This kind of survey reveals top items of concern, both good and bad, but is incomplete in the sense that different students may focus on different aspects and there is no mechanism to rank responses against each other. The "change" item encourages students to reflect on their own learning, which we consider a valuable use of their time, and it provides us with useful insight into their thinking. We grouped responses by theme for ease of interpretation and for planning changes in response to the feedback. In ASTR120, for which there were 40 respondents, we categorized the top 3 "stop" requests as: nothing (meaning nothing came to mind to stop, accounting for 27% of responses in this category); whiteboards/group work (15%); and going off topic (12%). The top 3 "go" requests were: instructors (meaning the students had a favorable view of the instructor and/or TA(s), accounting for 22% of responses in this category); participation/class discussion/peers (18%); and both voting questions (TPS) and whiteboards/group work (tied for 15% apiece). (Those listing whiteboards/group work in the "go" category outnumbered those listing it in the stop category by 2 to 1. We were still in the process of designing a consistent system for group work in ASTR120. It is notable this did not continue to be a concern for students in ASTR121.) The top 3 "change" goals were: reading habits (30% of responses); homework habits (11%); and note taking (11%). The emphasis on reading habits may stem from students' awareness that the redesigned course puts increased responsibility for preparing before lecture onto students. We also asked the students to identify one aspect related to their learning that they felt they were doing well; the top 3 responses were: reading (40%); assignments (27%); and exams (13%). Here it seems those students who were doing the reading felt it was beneficial for their performance in the course. Table 6.4 collects student responses that illustrate the stop-go-change results, grouped by the most prevalent categories.

Table 6.4: Selected student responses from the ASTR120 Fall 2015 mid-semester stop-go-change survey grouped by the most prevalent categories of responses.

STOP	
Nothing	"Everything about this class is actually pretty good so far!"
Whiteboards / group work	"I do not like the whiteboard problems. The whiteboards are cumbersome and tedious." / "I think our groups each class should be randomly selected for us so we don't work with the same people over and over."
Going off topic	"We get off topic a lot because people with more knowledge about the subject ask complicated questions with long answers."
GO	
Instructors	"Instructor does an amazing job of encouraging thought and participation."
Participation / class discussion / peers	"The class format is engaging and keeps me paying attention." / "The combined lecture and discussion-with-peers format is fun for me." / "I like working with my peers."
Voting questions (TPS)	"Voting questions are a great way to encourage critical thinking."
Whiteboards / group work	"I also like doing the board activities to wrap up a lecture." / "I like that we do the mini group projects because you get to see what others are thinking."
CHANGE *(how students think they can improve their learning in ASTR120)*	
Reading habits	"I should be reading the material more carefully before class, instead of skimming it to answer the MasteringAstronomy questions."
Homework habits	"I could be doing the homework more thoroughly. I feel like I really rush through it without trying to understand anything."
Note taking	"Take more notes outside of class."

A few interesting responses bear mentioning even though they did not fall in a top-3 topic. One student wrote, in the stop category, "I don't like that other students make me feel intimidated, since I'm an underclassman, so I'm afraid to speak up as much." This is noteworthy to us because we aim to reduce the intimidation that students may feel when asked to speak up in class through group work. While the majority of students wrote favorably about working with their peers, this comment indicates to us that further refinements to facilitating, framing, and/or structuring group work may be worthwhile. A related comment from a student, in the “change” category, was, “I’m still learning to work well in a group because I usually work on my own since I process information a little slower than others.” This suggests the student feels that group work is important but challenging, and more fruitful collaboration may require a shift in their thinking or behavior. We find this more encouraging than the previous comment: we want students to learn to work well with their peers, even if they are initially struggling to do so. However, the hint that this student may feel left behind by their peers could be a problem we could also work to ameliorate through instruction in later semesters. Finally, some notable comments in the “go” category related to student study habits included, “I also like the short quizzes because it makes me read the chapter thoroughly and because it is short/low point, it is not very stressful,” and “I like the daily learning goals because it helps keep me on pace/lets me know when I need to study more.”

6.3.2.2 ASTR121 Spring 2016

In ASTR121, we distributed a similar stop-go-change mid-semester survey for Spring 2016, with 32 respondents. In this case, we modified the “change” question such that students were asked to indicate one thing that improved about their own learning in this course sequence since ASTR120 (where all students had taken ASTR120 in Fall 2015). The top 3 “stop” requests were: nothing (21%); whiteboard uploads (15%); and assignment deadlines/time constraints (in discussion, 12%). These second two comments point to places where negotiation and refinement in course logistics may be necessary, but do not point to substantive problems in the course transformation efforts. The top 3 “go” requests were: instructors (18%); format/variety (of lecture, 15%); and atmosphere/content (13%; adjectives used included, “interesting,” “engaging,” “fun,” “challenging,” and “cool”). In the “change” category, students identified things they were doing better since ASTR120—reading/studying (50%); note taking (15%); and homework (15%)—and things they wanted to continue to improve: reading more/more carefully (31%); practicing more (studying, doing exercises in the textbook, and clarifying muddiest points, 23%); and taking more notes (12%). Again the emphasis on reading seems to reflect the central role this plays in the redesigned course, and students feel they need to do well with this aspect in order to succeed. Examples of student responses that illustrate these themes are given in Table 6.5.

Table 6.5: Selected student responses from the ASTR121 Spring 2016 mid-semester stop-go-change survey grouped by the most prevalent categories of responses. In contrast to Table 6.4 where “change” indicates what students aspire to improve about their participation in the course, this change section shows what students felt they were doing better since ASTR120 (all students in Spring 2016 took ASTR120 in Fall 2015).

	STOP
Nothing	“None. I really like the class format.”
Whiteboard uploads	“The whiteboards are a good idea, but the way in which we submit them is still a bit inconvenient.”
Assignment deadlines / time constraints	“I would like it if we could be able to turn in discussion activities a week later if we do not finish in class.”
	GO
Instructors	“I think the instructors do a good job.”

Format / variety of lecture	"I like how the class is consistent with how it deals out assignments in a regular and periodic fashion." / "I think the different components work well together to help me understand the material."
Atmosphere / content	"Engaging and fun." / "This class is interesting." / "The material is challenging and engaging while still always being clear." / "The material this semester is really cool." / "It's the most interesting and well-run class I am taking."
CHANGE *(how students think they have improved since ASTR120)*	
Reading / studying	"Since ASTR120, I have started making sure to do all of the readings before class. It has helped a lot."
Note taking	"I'm more active with the readings and I take notes both during my reading and in lecture."
Homework	"I'm also starting homework sooner."

As in the previous subsection, there were a few other interesting responses for the ASTR121 stop-go-change survey that we think add insight into the course redesign. In the "go" category, one student wrote, "Retests [allow] you to go back and make you learn what you missed on the test [and] also helps people with test anxiety," showing that the retests are being appreciated for their intended purpose. As in ASTR120, we also identified feedback that may point to challenges in group work. In the "change" category, one student commented, "I should probably try to get more involved with my group members but it's difficult to focus when I do that." This suggests to us that this student is struggling to collaborate smoothly with their peers; however, we are again encouraged that this student seems aware that becoming a more productive group member presents an opportunity for personal growth, even if they may need additional support or encouragement from the instructor in order to make these adjustments in their behavior. There was also a generally expressed desire to do even more sample problems in class, which shows that the students value the practice they get with hands-on work during lecture.

6.3.3 ASTR121 lab outcomes

6.3.3.1 E-CLASS data

For the lab portion of ASTR121, we used a modified version of the Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS) [Zwickl et al., 2012] administered at the start and end of the spring semester. The published, research-validated survey is designed to assess shifts in students beliefs and attitudes toward laboratory science as a result of their coursework, specifically, undergraduate labs in experimental physics. While a majority of the E-CLASS items seemed relevant for the ASTR121 lab, we removed a few items that pertained to the use of physical equipment and thus seemed poorly matched to our context (which is focused on the exploration of astronomical data using MATLAB, rather than carrying out physical experiments), and replaced physics with astronomy throughout. (We also note that we mistakenly used an earlier version of the E-CLASS with slightly different items when we distributed the E-CLASS initially, and have maintained these items in previous iterations for consistency.) While formal analysis of our full dataset is pending, we will discuss some preliminary results here.

Overall, students' beliefs and attitudes did not change in a significant way in either semester, but certain items did increase and decrease in interesting ways. In particular, the 2015 surveys showed that students' attitudes toward group work and collaboration in particular became significantly more expert-like over the course of the semester (item 15 in Figure 6.4; also see Table 6.6), the biggest such change in the items considered. This trend continued in 2016, with the notable difference that the positive attitude toward group work at the start of the semester was much higher in 2016 than in 2015 (85% compared to 50%), and as a result we saw a smaller change in expert-like response compared to 2015. While it is possible that the students in Spring 2016 simply started the course sequence with more expert-like beliefs about group work, it is also plausible to us that the greater emphasis on group work in ASTR120 Fall 2015 compared to Fall 2014 already influenced their responses, so students began the ASTR121 Spring 2016 lab with an already enhanced appreciation of the value of teamwork in an academic astronomy setting.

Table 6.6: List of adapted E-CLASS items used in Figure 6.4.

#	*Statement*
1	When doing an astronomy experiment, I don't think much about sources of systematic error.
2	It is helpful to understand the assumptions that go into the theoretical predictions when comparing them to data.
3	Doing error analysis (such as calculating the propagated error) usually helps me understand my results.
4	If I don't have clear directions for analyzing data, I am not sure how to choose an appropriate analysis method.
5	I am usually able to complete an experiment without understanding the equations and physics ideas that describe the system I am investigating.
6	I try to understand the theoretical equations provided in the lab guide.
7	Computers are helpful for plotting and analyzing data.
8	When I am doing an astronomy experiment, I try to make predictions to see if my results are reasonable.
9	When doing an experiment I usually think up my own questions to investigate.
10	When doing an experiment, I just follow the instructions without thinking about their purpose.
11	When I encounter difficulties in the lab, my first step is to ask an expert, like the instructor.
12	A common approach for fixing a problem with a lab is to randomly change things until the problem goes away.

#	*Statement*
13	Communicating scientific results to peers is a valuable part of doing astronomy experiments.
14	I am able to read a scientific journal article for understanding.
15	Working in a group is an important part of doing astronomy experiments.
16	If I am writing a lab report, my main goal is to make conclusions based on my data using scientific reasoning.
17	If I am writing a lab report, my main goal is to create a report with the correct sections and formatting.
18	The primary purpose of doing an astronomy experiment is to confirm previously known results.

6.3.3.2 ASTR121 lab stop-go-change

We also conducted a stop-go-change mid-semester evaluation for the lab (see Table 6.7; here the "change" question was similar to what was asked in ASTR120, namely, how might you improve your success in lab, and what are you doing well), for which there were 36 respondents. The top "stop" requests were: MATLAB (the programing environment used for all the labs, 18% of responses); wording/material in lab handouts (16%); and harsh/unclear grading (14%). The first two items seem to primarily be evidence of student resistance to aspects of the lab that, to us, are pedagogically valuable, namely, the limited explicit instruction and somewhat open-ended questions that require students to explore on their own and make judgments about what procedures and visual outputs would appropriately answer the driving question in the lab. This pushback may simply be a natural part of students adjusting to this instructional model, and it underscores the work that our instructors need to do throughout the semester to help students see the value in having increased agency in the classroom. For the third item, we do provide a detailed grading rubric

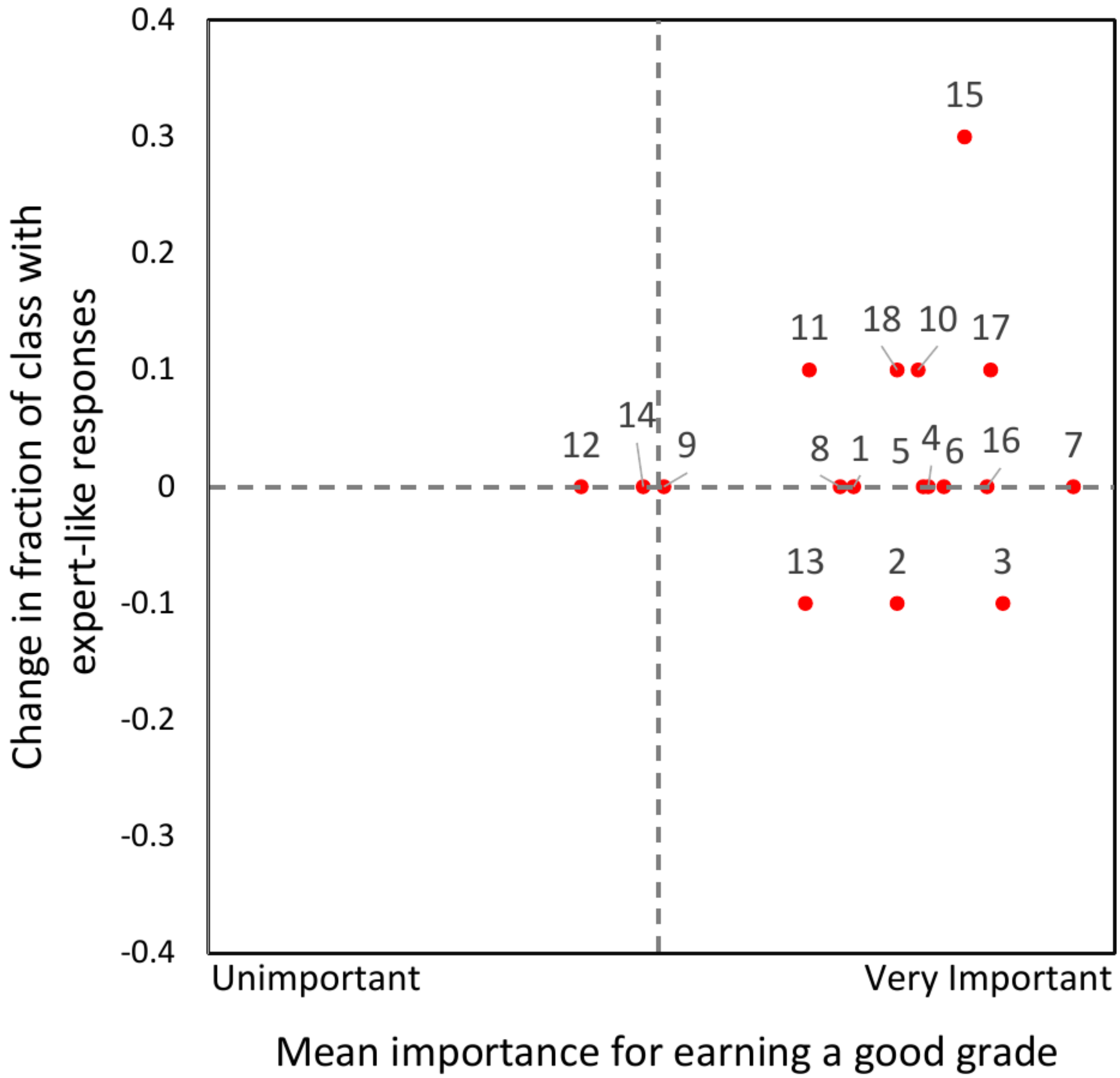

Figure 6.4: Plot of changes in expert-like views between the beginning and end of the ASTR121 Spring 2015 lab for our E-CLASS survey questions (numbered 1 to 18 here), as a function of the degree to which each is deemed important for earning a good grade in the lab. The horizontal coordinate for each point is a mean value. The vertical coordinate is the change in the fraction of the class with expert-like responses between the pre- and post-surveys. Most points that showed a change are in the upper-right quadrant, and represent items the students considered very important for earning a good grade and that showed a positive change in expert-like views. Item 15, which was the strongest in this respect, was worded as, Working in a group is an important part of doing astronomy experiments. This plot is modeled after the E-CLASS instructor reports [Wilcox et al., 2016].

for the lab reports, but most of the criticism was directed at the prelab exercises, for which there is no rubric. The top "go" comments were: instructor (in this case the lab TA, 21%); material (that is, the lab contents and procedures, 19%); and collaboration (18%; many students expressed appreciation about having a partner). Although not in the top 3, a number of students commented that they found the scaffolding approach to be very helpful (14% of responses). The comments regarding group work corroborate the E-CLASS data discussed above. The top "change" ideas were: work on MATLAB (a common theme for those who felt they needed more practice, 31%); read lab handouts (so students come to lab better prepared, which is part of the point of the prelab exercises but it appears students focus more on the questions than on the reading, 14%); and nothing/not sure (14%).

Table 6.7: Selected student responses from the ASTR121 Spring 2016 mid-semester stop-go-change survey for the lab, grouped by the most prevalent categories of responses.

	STOP
MATLAB	"How long it takes to figure out basic things in MATLAB on our own." / "I don't like how we have to learn MATLAB on our own in order to finish the labs. [I] wish there were some extra practice exercises to get used to MATLAB."
Wording/material in lab handouts	"Prelab questions often vague or not found in lab handout." / "Not enough background information, especially for the prelab questions." / "The instructions for the lab are often vague, and sometimes it is not indicated exactly [what] we are required to create."
Harsh / unclear grading	"Unclear grading for some parts." / "Harsh grading on prelab."
	GO
Instructor	"Instructor walking around to answer questions."
Material	"Feels satisfying when the code actually works." / "I think the lab content itself is interesting."
Collaboration	"I like the data in the labs and working with a partner or small group to solve problems in the various labs."
	CHANGE *(how students think they can improve their learning in the lab)*
Work on MATLAB	"I want to put in more time into learning MATLAB but it's tough with other homework."
Read lab handouts	"I could read the lab more carefully before class."
Nothing / not sure	"None." / "Not really sure."

6.3.4 Formal student evaluations

In addition to the insights gained from soliciting student feedback ourselves, we find that more standard student evaluations for ASTR120/121 have remained high, even increased, since beginning the course transformation process. For ASTR120, which was taught by the same instructor (Derek) in Fall 2014 (pre-redesign) and in Fall 2015 (post-redesign), the overall score for categories related to content rose modestly from 3.69 out of a possible 4 in 2014 (42 respondents) to 3.72 in 2015 (40 respondents), while the overall score related to the instructor rose from 3.77 to 3.93. As a whole these results indicate overall satisfaction and student buy-in to the redesign. Because our local teaching and learning center, the TLTC, funded this course transformation effort, they also distributed a 4-question, Likert-scale survey in Fall 2015 to evaluate our students' perceptions of their own learning and engagement in the course, as well as the instructor's attitude toward their learning. This survey also showed favorable students attitudes towards the course and the instructor overall (mean above 4.5 out of 5 in all categories, with 45 respondents). (The course evaluation and TLTC survey results for ASTR121 in Spring 2016 are not yet available as of this writing.) One student commented on the ASTR120 TLTC survey, "This class is absolutely fantastic, definitely my favorite this semester. You can learn a very large amount of material very thoroughly by attending class and participating in the many, very engaging, very helpful, class activities. This is an awesome introductory class for astronomy, and I definitely intend on continuing with the major." Finally, as additional evidence that students had equally if not more positive attitudes towards the instructor than in previous years, we also note that Derek was awarded the College of Computer, Mathematical, and Natural Sciences Dean's Award for Excellence in Teaching based on student nominations received in Fall 2015.

6.4 The change process

We now turn to look at the course transformation from a different angle. While we consider it valuable to communicate the current structure of the course and the student outcomes we have measured, as we have done in the first half of this chapter, we consider it equally valuable to communicate how these changes occurred and what factors seemed to support or limit them. As we mentioned in Chapter 2, using the framework Henderson et al. [2012] developed to characterize change approaches in undergraduate STEM, we locate this effort in the "developing shared vision" quadrant: the outcomes of the instructional change we undertook were emergent and team-based. In other words, our approach to instructional design required multiple people to agree on shared products, and the final form of the instructional changes was not predetermined.

Broadly speaking, team-based approaches to instructional change are both highly promising in their potential to address the limitations of other change models: for instance, they allow instructors to develop instructional approaches that match their unique instructional contexts and might provide needed long-term supports, unlike one-way, individually-focused dissemination of instructional strategies

[Henderson et al., 2012]. However, shared vision approaches are also under-studied relative to other approaches and thus the affordances and challenges surrounding them are not well-understood [Henderson et al., 2012].

In particular, team-based approaches to change can vary widely in scope, and this variation is likely consequential to the immediate outcomes and long-term sustainability. In this section, we present a descriptive account of the change process for ASTR120/121 as a case study of one small-scale instructional change effort characterized by developing shared vision. While we have limited direct evidence to show what about the ASTR120/121 change process was consequential, we can use prior studies of other, larger-scale "developing shared vision" approaches to guide our intuitions as we decide what to highlight and draw attention to. Specifically, we find the following team-level and environmental factors that others have proposed as supports that enable successful team-based instructional change efforts relevant in our analysis in this section:

Team-level factors

- A team comprised of people with a diversity of pedagogical and disciplinary expertise, including...
  - people with extensive knowledge of education research findings who can translate these ideas for other instructors and put them in disciplinary contexts [Bouwma-Gearhart et al., 2014, Wieman et al., 2010, Chasteen et al., 2011, 2015]
  - undergraduate students who can help to mitigate potential student resistance to course changes [Foote et al., 2016] and provide insight into

student thinking [Chasteen et al., 2011]

  – instructors with prior exposure to and/or experience using active learning instructional approaches [Wieman et al., 2010]

- A lead instructor who recognizes and values many kinds of expertise [Bouwma-Gearhart et al., 2014]
- A lead instructor who is dissatisfied with student outcomes from prior teaching [Wieman et al., 2010]
- Interpersonal trust and social ties among the team [Bouwma-Gearhart et al., 2014]

Environmental factors

- Support from department leadership and other institutional administrators [Corbo et al., 2016, Wieman et al., 2010, Bouwma-Gearhart et al., 2014, Chasteen et al., 2011, 2015, Foote et al., 2016]
- Sufficient resources for the proposed instructional changes (funding, people, space) [Wieman et al., 2010, Chasteen et al., 2011, Foote et al., 2016]
- Change efforts align with existing features of department culture [Corbo et al., 2016]
- Broad faculty support and involvement [Wieman et al., 2010, Corbo et al., 2016, Chasteen et al., 2011, 2015]

We draw on these factors as we consider what contributed to producing and sustaining the shifts towards increased active learning that we have documented, as well as how the sustainability of these changes could be limited. While this was not intended to be a formal research study at the onset, and therefore we did not have sufficient support (time or funding) for extensive data collection during the change process, we are able to construct a limited retrospective account using artifacts such as field notes, written reflections on instruction, past meeting agendas, and emails, to substantiate our recollection of what occurred. We aim to show how we found footholds for instructional change within our local context by illuminating salient features in the change process, and in doing so, hope to help other instructors to think critically about their local contexts and to help administrators consider how putting in place various environmental supports might enable similar change processes to occur.

6.4.1 Deciding to pursue ambitious changes

In this section, we describe the factors that contributed to and supported our decision to pursue a formal course transformation effort. We show a full timeline of events spanning the entire project in Appendix D. In this section, we consider the events that led to the formal (i.e., funded) start of the project in January 2015. We first describe the history of collaboration between myself and the lead instructor (Derek), and then describe the factors that motivated us to pursue funding from our local teaching and learning center for this particular course sequence.

6.4.2 The history of collaboration between Derek and myself

Foundationally, our history of collaboration was critical to starting this project. Derek has been a strong advocate for my pursuit of Astronomy Education Research

(AER) in the Astronomy department since I first considered doing so, and currently serves as my thesis committee co-chair. The intellectual focus of this relationship contributed to an increase in Derek's knowledge of and motivation to try RBIS that led him to implement these strategies for the first time. Because our local faculty had never (to our knowledge) been asked to permit an AER thesis previously, the process of gaining approval for this thesis research was non-trivial. As I worked to construct arguments for why my research matters, including why increased adoption or adaptation of RBIS could benefit students, Derek, as my primary representative among the faculty, started to develop fluency in these arguments as well. As I became increasingly knowledgeable about education research while beginning my thesis research in earnest, I continued to share my ideas with Derek regularly, thus simultaneously contributing to Derek's learning about education research theory and improving my ability to organize and articulate these ideas in ways that are comprehensible to practitioners.

Derek and I had also established a shared orientation toward the process of teaching, where we both agreed that improvements to instruction are possible and attainable, and a high degree of trust around evaluating each other's instruction. In our department, as is true in much of academia/higher education, teaching is often viewed as a solo pursuit: faculty rarely observe each other's instruction, and may be uncomfortable being observed because they risk negative judgment from their peers [Hora and Ferrare, 2013, Chism, 2007]. In contrast with these norms, Derek and I have taught together and discussed instruction on many occasions, which has helped to reduce these barriers. In particular, we co-taught ASTR121 in Spring

2014, following a positive TA-lead instructor relationship during ASTR120 in Fall 2012, which was highly unusual for our department. During these shared experiences, Derek and I also shared instructional challenges and goals, and established norms that contributed to an ease in diagnosing common challenges in our later interactions. Furthermore, my developing knowledge of PAER prompted both of us to start experimenting with RBIS in our own classrooms. We both had positive experiences trying out TPS briefly while co-teaching ASTR121 together, and when we next taught separate, full-term classes in Fall 2014, both used TPS more extensively. In general, while I incorporated more student-centered, active learning strategies into ASTR100/101 in that semester overall, I frequently made suggestions and supported Derek in incorporating similar pieces into ASTR121 by sharing full strategies (e.g., Lecture-Tutorials), specific student tasks (e.g., TPS questions), and physical materials (e.g., voting cards). On top of these academic experiences, Derek and I established a friendly social relationship and would frequently meet up in a purely social capacity, which likely also contributed to smooth collaboration and trust heading into this formal project.

6.4.3 Identifying opportunity

When our local Teaching and Learning Transformation Center (TLTC) put out a call for proposals for course transformation projects in Fall 2014, we were well-prepared to see this as an opportunity. For one, Derek had access to what we felt was a strategic course. As an introductory sequence, ASTR120/121 could serve as lever to improve the experiences of astronomy majors, as improved preparation early on could set undergraduate majors on a productive trajectory for the duration of their time at Maryland. As mentioned in Section 6.2, we had also identified one obvious initial target for the reform effort—the ASTR121 lab—which we both recognized as especially problematic and potentially high-leverage for improving students' later success.

ASTR120/121 was also a strategic choice because it seemed to have the most potential to positively influence the instruction of other faculty. In part, we hoped that student buy-in to active learning strategies early on might trickle up to future courses, where students might suggest or request these pedagogical approaches of their astronomy instructors later on [Chasteen et al., 2015]. Similarly, there was the potential for other faculty to see benefits of our changes for these cohorts of students, since faculty regularly teach the same majors' courses multiple times, and thus become more interested in pursuing similar strategies themselves. Moreover, we thought that changes to the curriculum and pedagogy of majors' courses, particularly this foundational sequence, were more likely to be adapted or adopted by future instructors than changes to non-majors' courses: faculty seem much more likely to have similar views about what constitutes critical knowledge and skill development in ASTR120/121 as compared to a non-majors' "Astro 101" course, where faculty's goals are known to be highly variable [Slater et al., 2001].

Participating in the TLTC program had projected affordances that would enable these changes to take place. Successful proposals were provided with resources such as funds ($\sim$\$12K) that could be used to hire team members, opportunities for the ASTR120/121 lead instructors to participate in regular, guided discussions about course transformation with faculty from across campus, and access to guidance from TLTC staff. We also recognized that participation in the TLTC program could draw increased attention and prestige to our efforts within the Astronomy department, and allow us to gain recognition from university-level administrators, which might increase the likelihood that changes would be sustained beyond the duration of the grant.

Applying for the TLTC grant also provided opportunities for us to have conversations with our department chair about the intent behind these course changes, and to demonstrate alignment with other departmental initiatives. Engaging seriously with equity concerns and considering interventions that could increase the success of underrepresented groups within astronomy has become a focus of our department over the past few years. Two of these efforts—a program that aims to create research and networking opportunities for undergraduate students at minority-serving institutions and a women in astronomy mentoring group—were initiated by graduate students, but have also been strongly taken up and endorsed by our department chair. Our department chair has since initiated a seminar series (Better Astronomy for a New Generation or BANG!) that largely focuses on equity issues and an equity and inclusion committee to work on department-wide challenges, and has become a member of the American Astronomical Society Committee for the Status of Women in Astronomy. Because of this, we naturally found alignment with our department chair's values and concerns because we were also strongly motivated by the potential for these course changes to increase opportunities for success for students who tend to be the most disadvantaged by lecture. This alignment with departmental leadership has been an asset during the course transformation process, and allowed us to recruit specific team members more easily than might have been the case otherwise. We describe this recruitment process in the section that follows.

6.4.4 Assembling a team

In this section, we describe the diversity of expertise on our instructional development team and the ways in which we recruited team members. Derek and I were both well-prepared to take on this course transformation project, but felt that bringing in team members with complementary expertise would be highly valuable. In particular, we highlight ways that our recruitment of team members was both deliberate and tied to various structural environmental features of our university, such as the past hires of instructors committed to implementing (and/or willing to try) active learning strategies, the existence of a local LA program, and the recent attention to cultural change within the Astronomy department.

Derek and I recognized from the onset that other faculty would bring valuable instructional expertise and could lend unique stability to the course changes in the long term because of their potential to become lead instructors of this course sequence in future semesters. Unfortunately, our ability to recruit and keep faculty on our team has been somewhat limited so far. In particular, while we initially recruited a tenure-track faculty member who we hoped would teach ASTR120/121 following Derek, he left the University of Maryland for another faculty position soon after this project started, and thus was never highly involved in our course transformation efforts. We did successfully (and easily) recruit a local Astronomy lecturer, Dr. Melissa Hayes-Gehrke, though she has fewer opportunities to teach majors-level courses than the local tenure-track faculty and is unlikely to be given the opportunity to teach this course sequence for structural reasons. However, her instructional expertise allowed her to easily contribute to and advise the course transformation effort in substantive ways: she has participated in many national and local teaching workshops, has implemented a wide variety of research-based instructional strategies in her classes, typically "flipping" classes to allow more time for student discussion and debate, and has won multiple teaching awards at the University of Maryland. She had already been acting as an informal collaborator for us, and had suggested small instructional improvements that Derek had already implemented, making her an obvious choice as a formal collaborator.

In contrast to the small number of faculty involved, we recruited many undergraduate and graduate students to contribute to this project: in total, five undergraduate students and two graduate students (not including myself) participated in the project across three semesters, and were paid for their efforts. We felt strongly that well-prepared students can take substantive roles in developing course materials and give valuable feedback about how to engage students, therefore we invested time into recruiting these students strategically whenever possible. Some of this recruitment was second-wave: two of the undergraduates who participated (Sarah Scott and Fatima Abdurrahman) were LAs first, and thus had been recruited through a broader and more systematic application process in the past. Moreover, their participation in the LA program meant these undergraduate students had experience teaching in highly interactive classroom settings and had developed some expertise in pedagogy through formal coursework (a required 3-credit seminar). We also recruited undergraduate and graduate team members from our own former students, TAs, and LAs. From a pedagogical perspective, this meant that many team members already had some concrete understanding of each other's instructional vision prior to working on the course transformation effort, which made it easier to develop a shared understanding of potential changes. In particular, we were able to recruit an undergraduate (Joseph DeMartini) who experienced the first round of changes to ASTR121, which gives him valuable insight into our current students' experiences and makes him well-positioned to get students to buy-in to the reform. From a social perspective, this meant that we already had a rapport with many of these people, thus making us more confident that collaboration would be relatively smooth. In some cases, our recruits fell into multiple categories: for instance, Fatima had been a student of Derek, Melissa, and I at various points, as well as an LA for my Astro 101 course and the transformed introductory physics for life sciences course in the physics department [Redish et al., 2014]. More specifically, she had extensive experience teaching reformed labs and had recently taken a lab-intensive course from Melissa, making her highly qualified to take a lead role in revising the ASTR121 lab. We also sought to further train our team members by encouraging new members to take or audit the LA pedagogy course, which was possible in one case (Sara Frederick).

Less directly, the recent cultural push to focus more on equity issues within the Astronomy department positively influenced the individual expertise and sense of shared purpose within our team. In particular, even though we were not able to recruit and select all of our team members from within the Astronomy department, i.e., some TAs were assigned to this course sequence based on departmental scheduling constraints, we think that having well-established equity initiatives and advertising to prospective graduate students contributed to the department as a whole recruiting students who would likely be enthusiastic about making undergraduate students' experiences more equitable through this course transformation work. In particular, we think this department-level recruiting was at play in Fall 2015, where one of the incoming graduate TAs (Sara) was randomly assigned to ASTR120 and indeed has become both highly committed to this project and involved in other equity initiatives. We suspect that several of the other graduate TAs in her cohort would have been similarly committed to and enthusiastic about this work. Department-wide efforts to build community around making astronomy culture more inclusive have increased the sense of community within our team as well, as most team members have led and/or participated in these initiatives. Because of this, it is typically easy for our team to find common ground in considering how social science results like stereotype threat are relevant to student learning, and to think together about how to address this.

6.4.5 Team routines

In this section, we describe the set of routines that our team followed throughout the course transformation process. Specifically, we describe the typical structure to the team's collaboration during and outside of meetings, then define and elaborate on the typical activities enacted during meetings.

6.4.5.1 Structure of collaboration

Our team interacted in a variety of settings: weekly team meetings, virtually, in the classroom, and one-on-one meetings. Here, we briefly describe the frequency, type, and structure of interactions that were typical of our collaboration.

Our 1-hour weekly team meetings were a central interaction space. They include all active team members and have a flexible organizational structure. Derek wrote agenda items in a shared Google Doc ahead of time, which we displayed to help guide our time management during the meeting and keep track of any team tasks generated during meetings. This document contains a record of all of the meetings for a given semester. Derek color-coded previous action items and the person responsible for addressing that point, which enabled us to quickly revisit any outstanding team tasks at the start of the meeting and ask for progress updates. Some class-specific items were flagged as tasks to work on in the following year, allowing the team to focus on time-sensitive issues without becoming weighed down by attempting to perfect the course in the first round. We typically limited discussion of logistics that are only relevant to a few team members, such as collecting, grading, and returning papers, during whole team meetings, and Derek typically spent a few (3–5) minutes immediately after any particular meeting making detailed plans with the active TAs/LAs.

Regular team meetings enabled significant work to occur outside of team meetings while maintaining a unified vision for the course. Undergraduate and graduate team members (beyond the first author) were often delegated to write student tasks that had been initially brainstormed during team meetings, or to revise tasks based on the team's feedback. For example, undergraduate and graduate students frequently take lead roles on creating and revising lab assignments, discussion activities, and sometimes TPS questions. When we delegated work in this way, the team leaders (Derek, me, and sometimes Melissa) provided additional feedback and guidance via email, one-on-one meetings, and future team meetings as requested. This structure allowed us to make progress on multiple aspects of the course at once, and likely instilled a sense of ownership in our TAs/LAs that may have contributed to their investment in and attention to student outcomes. Similarly, email communication also provided space for Derek to solicit feedback on initial drafts of learning goals and begin to revise them.

Team members also regularly observed each other's instruction and students' engagement in classroom activities during the change process. Sometimes these opportunities arose because we structured the course to include co-facilitation of student group work, while at other times team members have intentionally observed each other's instruction to learn and give feedback. For example, all six team members who played a role in developing the lab component of ASTR121 attended a lab session at least once, often multiple times, as these changes were being enacted and subsequent labs were being designed. I also observed discussion sections and lectures in both Spring and Fall 2015, and offered one-on-one feedback to Derek and the TAs/LAs. As referenced in Section 6.1, I used the Classroom Observation Protocol for Undergraduate STEM (COPUS) to document the lecture portion of ASTR120/121, and debriefed with Derek using this data about the prevalence of specific kinds of interactions and behaviors to ground these conversations. Sara routinely took field notes on students' engagement during lecture (Fall 2015–Spring 2016) in addition to co-facilitating student group work. Derek also observed several other instructor's classes outside of ASTR120/121 to expand his vision for what active learning approaches can look like.

In addition to being observed by others, several team members took notes on their own instruction following class. These notes were sometimes primarily intended for internal reflection: for example, in Fall 2015, Derek and Sara both regularly wrote reflections on their own instruction and observations of student engagement in lecture and discussion, respectively. These written reflections acted as an asynchronous form of communication between team members: Derek's reflections built on Sara's detailed field notes, and I regularly read through these reflections to write comments and questions. These asynchronous communications were often synced in team meetings, where instructional challenges mentioned in these reflections were often brought up to the whole team. In other cases, written reflections were primarily intended to inform future instructors: several of our TAs/LAs took notes following class to communicate their instructional choices and the resulting successes and limitations that likely arose because of these choices. For example, Fatima wrote an instructor's guide for the redesigned lab in this way, and the Fall 2014–Spring 2015 TAs/LAs took notes on which discussion activities to keep, revise, or discard. All of these notes have been used subsequently to improve course materials and avoid potential pitfalls.

Lastly, one-on-one meetings between Derek and I allowed us to have more direct and/or critical conversations about the team's progress, Derek's instruction, and the use of our weekly meeting time. These interactions well-represent a way in which Derek and I had joint leadership over this project—I was able to make suggestions for agenda items or problematize other aspects of the course transformation without contesting Derek's leadership role during team meetings, and Derek almost always took up these suggestions and shifted the direction of the team's efforts accordingly.

6.4.5.2 Focus of collaboration

In the following paragraphs, we describe, justify and give examples of the kinds of activities that occurred within team meetings, supported by the general collaborative structure described above. Specifically, our group regularly used meeting time for: (1) defining or refining learning goals; (2) creating and refining student tasks; (3) debriefing about observations of student engagement; (4) interpreting assessment results; and (5) planning for team presentations and reports. Several of these activities were typically enacted within any given team meeting, and comprise an exhaustive list of what typically occurred.

Defining and refining learning goals: We devoted significant energy to defining learning goals for every class. As many education researchers have argued, we consider defining learning goals to be a foundational first step in designing all other assessments and tasks, and in making expectations transparent to students [Chasteen et al., 2011]. As mentioned previously, no class-specific, lab-specific or course-level learning goals had been articulated prior to the start of this project. In Spring 2015, we focused on creating learning goals for the redesigned lab and the course overall, and intentionally waited to start creating class-specific learning goals until the following fall, which aligned with our general redesign strategy to focus our efforts on the lab first. In the following year, when we began to make more changes to the lecture portion of the class, team members helped to align class-specific learning goals with the kinds of student reasoning that we want or expect them to develop, avoid unnecessary jargon, and improve the clarity and accuracy of science ideas. Some examples of proposed and revised learning goals are shown in Table 6.8.

Table 6.8: Example proposed and revised learning goals from ASTR120, and an explanation of the intended effect of the changes. Italicized text in the learning goals indicates changes. All learning goals take the form "At the end of this class, students will be able to…"

Proposed learning goal	*Revised learning goal*	*Intent of changes*
…use Kepler's laws to describe the motion of solar system bodies and *graph the relevant quantities*.	…use Kepler's laws *and the properties of ellipses* to describe the motion of solar system bodies, and *graphically represent each law*.	Improved scientific accuracy; better alignment with what students are really doing; more specific.

Continued on next page…

Proposed learning goal	*Revised learning goal*	*Intent of changes*
. . . demonstrate why the Earth exhibits tides *(with two high tides and two low tides every day)* and tide seasons *(monthly neap and spring tides)*	. . . demonstrate *visually and mathematically* why the Earth exhibits tides, and *how they change throughout the month, and apply the underlying concepts to other similar astronomical situations.*	More specific about what students are doing; reduced jargon; connects to similar astrophysical situations relevant for this class.
. . . *explain how planetary interiors gain and lose heat, and how this is related to geological activity, including the presence or absence of magnetic fields*	. . . *predict the interior structure and geological activity of a terrestrial planet, and whether the planet has a magnetic field, based on the planet's size, spin, distance from its star, and age*	Separates underlying physical properties from characteristics that students are expected to infer.

Creating and refining student tasks: Our team often considered tasks that would be given to students during our team meetings. These tasks took a variety of forms: lab assignments, TPS questions, discussion activities (including Lecture-Tutorials), and whiteboard activities. We note that these tasks were all "new" to the course structure, and with the exception of pre-lab assignments, students worked on them during class; we rarely created or refined familiar tasks like homework assignments or exams as a team. Again, this was a matter of prioritizing: Derek already had a familiar routine for writing homework assignments and exams, and was easily able to make small revisions to his design process to better align these tasks with the newly defined learning goals. In contrast, the other activities needed to be newly envisioned and were the most directly linked to students' engagement in class, making them critical for the whole team to discuss. We often brainstormed these new activities together during meetings, further developed ideas that had been proposed over email, and gave feedback on drafts that team members (TAs, LAs, or Derek) had created. Brainstorming was particularly productive because our team had a breadth of individual and shared experiences in teaching and learning astronomy, and we could often build on these initial ideas to generate new tasks. This strategy prevented Derek (and other team members) from getting "stuck" in finding appropriate student tasks—a challenge that other physics faculty have referenced as a reason not to persist in using TPS/Peer Instruction [Turpen et al., 2016]. The improved quality of the tasks that resulted from team collaboration likely led to a greater chance of supporting desirable student behaviors in class, thus helping to sustain reforms in multiple ways [Turpen et al., 2016]. We revisit the potential connection between success at engaging students with these tasks and sustaining this reform effort in Section 6.4.7.

The value of team members' varied expertise was particularly noticeable when creating and refining new lab materials. Team interactions during meetings led to fine-grained improvements to labs that strengthened alignment with what we know to be valuable from education research. Specifically, these changes were often aimed at adding prompts that further encourage student collaboration, creating more space for students to explore and invent specific approaches to completing the lab by removing excess exposition, and re-framing of the lab to better support students sense-making (e.g., better defining the purpose of students' activities).

Debriefing about observations of student engagement: While developing learning goals and activities for the upcoming week was typically a high priority, our team also considered how these activities play out in class. Team members were able to bring up challenges in getting students engaged that they were unsure how to address, as well as successes that bolstered the team's enthusiasm and reveal particularly fruitful aspects of instructional changes. The high frequency of team members observing each other's instruction likely contributed to the productivity of these conversations by enabling instructors to corroborate and expand on each other's understanding of how instruction played out that week. As we argue extensively in Chapter 4 of this thesis, prior research has shown the importance of drawing on concrete classroom experiences in planning next steps [Aubusson et al., 2010, Horn, 2010, Morrell and Schepige, 2012, van Es and Sherin, 2008, Sherin and van Es, 2008], and we can see ways in which this worked well here. Bringing up ongoing challenges allowed the team to think collectively about what pedagogical choices could improve student engagement, and noting successes helped us hypothesize about what characteristics of the task design or implementation might be productive to repeat in later classes. For example, when students have resisted group work, we discussed strategies for encouraging more student-student collaboration such as requiring students to change seats and framing collaboration as important for students' success, and we later debriefed on our perceptions of the effect of implementing these instructional moves.

Interpreting assessment results: When our team used formative or summative assessment measures, we would often plan the distribution/form of these surveys and discuss interpretations of results during team meetings. The most central example of this is our stop-go-change mid-semester student evaluations, which we have distributed every semester. A team member, typically a TA or LA, would go through student responses, and roughly code and tally the results before the meeting (as evidenced earlier in Section 6.3.2). The team then discussed any successes and challenges, and determined next steps both in terms of deciding what modifications to the course are appropriate and considering how to debrief the results with students. For example, Sara's field notes from lecture show that Derek used the team's synthesis of stop-go-change responses to both demonstrate to students that their peers are satisfied overall, and that he would try to be responsive to their most prevalent concerns:

> *"[Derek] went through [the] evaluations, showed all responses but emphasized most popular stops-gos-changes, and where positive responses outweighed negative in e.g. TPS questions, group work. Derek mentioned that he would adjust the class in the future for certain suggestions e.g. Muddiest Points, and mentioned that students could email more suggestions."* November 8, 2015

In addition to these open-response assessments, we have implemented several research-based assessments, which we intend to use to justify and track the improvements to this course sequence as we have begun to do in Section 6.3. In contrast with some common assumptions about the use of data in the change process, we note that this summative data was not a strong driver for future changes. However, as others have noted, we do consider data to be useful after the fact to present on what we did, and suspect it could have value in communicating with administrators in the future [Henderson et al., 2014, Foote et al., 2016].

Planning for team presentations and reports: Our connections to the TLTC, the education research community, and the leadership in the Astronomy department, led to several opportunities for team members to share our work at meetings and conferences. We sometimes prepared and gave feedback on these presentations and reports during team meetings. These opportunities were generally productive for the team because they required us to synthesize our progress and help us to recognize the value in what we have done. For example, creating a poster for a regional LA conference gave Fatima the opportunity to identify pedagogical themes that spanned her previous three semesters of LA experience, consider how these themes were supported by literature she read during the LA pedagogy course, and describe how they played out in revising and implementing the ASTR121 lab. Figure 6.5 shows an excerpt from her poster.

6.4.6 The TLTC faculty learning community

As part of the TLTC Elevate Fellows program, the team leaders (Derek, Melissa, and I) also attended regular meetings with a cohort of other instructors, i.e., a faculty learning community, biweekly in Spring 2015 and monthly (with a larger group) in Fall 2015. We also had occasional consultations with TLTC staff over the summer. Melissa and I came into these meetings with a high level of knowledge about the research-based instructional strategies being discussed and demonstrated, and Derek had more experience than most of the other participants by virtue of his collaboration with Melissa and me. Nonetheless, Derek reports that he found the meetings affirming and they exposed him to a larger variety of practices used in other disciplines. Examples and scenarios presented in the meetings were interac-

Figure 6.5: An excerpt from Fatima's poster about her experiences as an LA within ASTR121, and how her previous experiences as an LA in labs for PHYS131/132—a transformed physics for life science majors course [Redish et al., 2014]—and ASTR101—introductory astronomy for non-majors—informed her instruction.

tive, which gave Derek a needed perspective on how these elements actually play out with students. It was also gratifying to our team to see the close correspondence between TLTC recommendations and the course design elements we already had in place, and required milestones (reports, course learning goals, etc.) helped keep our own design strategy on track. Lastly, these meetings and interactions with the faculty learning community gave many opportunities to share and present our redesign to other faculty and campus leaders. These opportunities included informally sharing with other members of our faculty learning community during regular meetings, a peer observation exchange with another TLTC faculty fellow, and formal presentations and social gatherings with members of the broader TLTC community, and helped Derek to recognize himself and be recognized by others as having expertise in this area.

6.4.7 Salient changes to Derek's instruction

As is true for most faculty learning communities, we suspect that a central goal of the TLTC program was to help Derek become more reflective on his instruction and thus support him in making positive changes [Henderson et al., 2012, Beach and Cox, 2009]. While we think that all of the members of our team learned from participating in this process, we find it interesting to briefly consider changes to Derek's instruction because of his role as a tenured faculty member in the Astronomy department, and his sustained leadership role in all parts of this course transformation effort. Furthermore, we think it is valuable to point out that instructional change is not instantaneous, and recognize that there are many possible trajectories for changes both towards and away from active, student-centered instruction. Derek's written reflections after each lecture, in addition to Sara's field notes and my occasional written commentary, provide evidence to help us reconstruct some shifts that we did not plan for initially and identify potentially salient factors that supported Derek's persistence in using new instructional strategies when problems arose. In doing so, we find evidence for the following:

– Establishing new classroom routines that fit with both Derek's instructional preferences and created space for these new reforms took time and experimentation.

- A lack of these routines early on posed a threat to Derek's persistence.
- Derek's initial successes in getting students engaged likely contributed to his initial persistence, where the primary ways he perceived "success" were tied to aspects of student engagement such as hearing novel student ideas, student enthusiasm, and variety in student participation.
- The quality of student tasks produced by our team seemed to contribute strongly to these successes.
- Establishing and obtaining evidence of student buy-in for the course format overall also seemed to contribute to Derek's persistence. Conversely, a lack of student buy-in may have posed a threat, both based on Derek's apprehension prior to obtaining feedback and broader trends documented in the literature [Turpen et al., 2016].
- Growing dissatisfaction with student engagement around certain structural elements of the course (deriving equations at the board) led to adding unanticipated instructional strategies.
- Observations of team members who observed Derek's class strengthened and motivated productive instructional changes.
- The incorporation of active learning strategies within Derek's instruction seems to have stabilized once he found new classroom routines that worked for him.

Each of these factors seems to have centrally shaped Derek's experience in this course transformation effort, and ultimately led to his increased comfort in using active learning strategies and a high likelihood that he will persist in doing so in the future.

6.4.8 Threats and assets to continuation

While we have primarily highlighted supports and successes in the course transformation process, there are limitations to our work from an instructional change perspective. In particular, while we hope that many of the instructional strategies and materials we have incorporated into ASTR120/121 will be adapted, refined, and built on by future instructors, there are numerous reasons why this may not occur. Here, we first outline these potential threats to continuation, and then consider factors that could ameliorate these challenges.

6.4.8.1 Threats to continuation

One significant threat to continuation is tied to who has participated in shaping the course, and more importantly, who has not. We suspect that other faculty in the Astronomy department perceive ownership of the changes as limited to current team members (Derek, Melissa), which is reasonable given the way the change process unfolded. Because Derek will not teach ASTR120/121 indefinitely, it is unclear whether any aspects of the changes we have made will be considered and built on by future instructors, or if most or all of these changes will be ignored and discarded. Few faculty in our department have extensive experience implementing these kinds of instructional strategies, and we have only limited understanding of individual faculty's interest in trying out different instructional approaches or what might motivate them to do so.

The degree of success of future implementations of the course is also contingent on the survival of the LA program: pedagogically trained undergraduates and graduate students have been a strong asset in facilitating student group work in this interactive classroom environment, and the future of the Maryland LA program is uncertain. Similarly, administrative support has been critical in bringing LA, TA, and grader support into the classroom, and will become more important as this course sequence no longer has direct financial support from the TLTC. Administrators can also work to enable the selection of thoughtful instructors to work in this course context, and have done so up until this point. Because of this, administrator turnover could also become a challenge: there is often limited institutional memory when people in these positions change, and new administrators may have different priorities.

6.4.8.2 Assets for continuation

Despite these potential challenges, there are some factors that could ameliorate them. While the ownership of the course transformation has been largely confined to team members, we have gained some recognition within the Astronomy department more broadly. All faculty in the Astronomy department are aware that these course transformation efforts are underway, and several have expressed interest in learning more about these changes. Derek presented on ASTR120/121 at a department-wide seminar in Fall 2015 and a faculty meeting in Spring 2016, and his extensive efforts were recognized through a departmental teaching award around that same time. We have gathered evidence of alignment with departmental needs by gathering informal feedback on what changes were needed from both faculty (via email and in-person discussion) and former students (through an in-class discussion with ~20 students in a subsequent course), which might lend weight to the changes we have made. For instance, a majority of former students' comments corroborated the challenges we identified in the previous lab curriculum, while students who have experienced the transformed course typically perceive the course format quite favorably. Recent students' first-hand accounts of their experiences in the course, along with their successes in future coursework or research that may correspond with being better prepared early on, could demonstrate the potential value of various course changes to other faculty.

As alluded to previously, we note that the lab seems like the most sustainable component of the reform. While the disconnect between the lab component of ASTR121 and the rest of the course has likely been detrimental to students' experiences in the past, this disconnect largely arises from past instructors' tendencies to leave the lab unaltered or asking TAs to improve on previous versions rather than trying out an entirely new lab curriculum. Thus, it seems reasonable to think that future instructors will be willing to leave the reformed lab curriculum largely unaltered, even if they decide to teach the lecture portion of the class differently. Recruiting former students to act as TAs/LAs and providing them with our existing, detailed lab instructor's guide along with in-person advice about negotiating student buy-in seems like a promising strategy for facilitating the lab, and within the next few years, almost all of the potential undergraduate Astronomy TAs will have experienced this transformed lab curriculum as students. Graduate TAs would also benefit from the extensive documentation we have put in place over the past two years.

If other faculty do decide to build from the other changes our team has made, they could be well-supported in a variety of ways. As mentioned previously, we have significant rapport with current administrators at the department and university levels, which would likely translate to future instructors. Our LA program director has strengthened ties with the Astronomy department leadership as a result of these sustained interactions over the past two years, and it is possible that LAs could be placed in this course should the LA program become institutionalized. All of the course materials we created are well-documented and well-organized, including lab assignments, instructors' guides, learning goals, whiteboard activities, and other tasks designed for students. Derek is scheduled to teach this course sequence for another year, and thus will be able to further improve and organize these materials if other instructors decide to build from them. More generally, team members are highly invested in the changes made to this course sequence, and many will remain in the department for several years if not more. While this could cause tension with future instructors, there is also the potential for collaboration around course changes to continue when other faculty teach this course sequence. All team members have both a strong understanding of why changes were made and in what challenges remain, and their continued involvement could create opportunities for other faculty to refine the course material in highly productive ways that match their own instructional expertise.

6.5 Conclusions

In this chapter, we described how we transformed a two-semester introductory astrophysics course for majors at the University of Maryland, ASTR120/121, with the goal of incorporating research-based instructional strategies to improve outcomes for all students, and particularly for students who are systematically disadvantaged in traditional lecture-based classes (i.e., women and/or underrepresented minority students). We established learning goals for the lectures and labs, and we aligned assignments and assessments with these learning goals. In the past year of this course transformation, typical classes began with Think-Pair-Share (TPS, or Peer Instruction) questions based on the pre-assigned reading, and the responses to these questions and muddiest point surveys were used to adjust the emphasis of the lecture material. Time was set aside for at least one whiteboard activity per lecture in which groups of students work on a problem together. These components taken together increased student engagement during class beyond what one would expect in a traditional lecture, as corroborated by classroom observation data. To further support student learning in the course, we provided videos of mathematical derivations, allow students to do optional midterm group retests, and arrange for out-of-class tutoring. The weekly discussion was tailored to the learning goals and often used research-based Lecture-Tutorials activities, with augmented content to increase the level of difficulty and/or incorporate more mathematically-oriented tasks. We also redesigned many aspects of the ASTR121 lab, giving more time for fewer projects, providing less explicit instructions to give students more freedom to explore, assigning pre-lab readings, using peer reviews to allow students to critique their work, and employing faded scaffolding (a lab report template that becomes increasingly sparse during the semester) to support students in learning to write high-quality lab reports.

We have identified some promising signs in the student outcomes that resulted from this course transformation. First, we note that students' conceptual learning following ASTR120, as measured by the Light and Spectroscopy Concept Inventory (LSCI), significantly increased in both Fall 2014 and Fall 2015. While the more dramatic instructional changes made in Fall 2015 relative to Fall 2014 did not seem to have a measurable effect on students' conceptual learning, we note that the effect sizes were high in all semesters considered (Fall 2014, Fall 2015, and Spring 2016), and the changes to the course structure did not degrade students' learning from year one to year two of the course transformation. Furthermore, while we lack comparison data from other astronomy majors' courses, we note that students' normalized learning gains on the LSCI following instruction in ASTR120 were comparable to gains in the most successful classes in a large, national study of student learning in Astro 101 [Prather and Brissenden, 2008] despite much higher pre-class scores, indicating that our astronomy majors left ASTR120 (and ASTR121) with significantly more knowledge about these topics than Astro 101 students following instruction. Thus, while it is unclear how much students would have learned about light and spectroscopy in an entirely "traditional" version of ASTR120/121, we do see evidence that experiencing this transformed course sequence significantly improved students' understanding of these astronomy key concepts.

Students' perceptions of their experiences in ASTR120/121 during the course transformation were also largely positive, as evidenced by open-response mid-semester evaluations, a research-based assessment instrument, and university evaluations. In particular, many students perceived value in the primary changes made to the lecture portion of the class, including TPS questions, whiteboard activities, and frequent discussions with their peers. There was some resistance to the transformed ASTR121 lab, which is not particularly surprising to us given the large responsibility placed on students as a result of the redesign. However, many students also saw the elements of the transformed lab favorably and recognized the affordances of our instructional approach. For example, data from a modified research-based assessment instrument, the Colorado Learning Attitudes About Science Survey for Experimental Physics (E-CLASS), collected in Spring 2015 and Spring 2016 showed a strong correlation between expert-like response for the importance of group work in astronomy research and earning a good grade in the lab. This suggests that students saw the prevalent collaboration in lab as contributing to their ability to successfully complete lab assignments, even though their grades were only directly tied to the written work they submitted (i.e., there was no participation grade and no formal attempt to track the extent of students' collaboration). This data also demonstrates that students left the course with favorable attitudes towards group work, which was encouraged across this course sequence. University-administered student evaluations remained as high as (potentially higher than) in previous semesters, indicating that while some student resistance was present, student buy-in to the transformed course sequence was strong overall. At the same time, some student comments regarding working with their peers suggest that further refinement in how group work is structured, framed, and facilitated in both the main part of the course and the lab could be beneficial, and merits attention in the future.

From an instructional change perspective, several factors seemed to contribute to the successes of this course transformation effort. In particular, we suggest that key supports included the diversity of expertise on our instructional team, including: team members with extensive training in education research and pedagogy; regular, well-structured team meetings focused on specific problems of practice and planning for future instruction in substantive ways; the collaborative development of course materials including the delegation of specific tasks to TAs and LAs; a department chair who was supportive of both this course transformation effort and broader equity initiatives within our department; and sufficient resources to compensate team members for their efforts. These factors align with other researchers' suggestions within the instructional change literature. At the same time, viewing this course transformation process from this broad perspective also reveals limitations in our work. In particular, we note that other faculty's involvement in this change effort was extremely limited, and largely as a result, the sustainability of these instructional changes within our department is highly uncertain. While our retrospective analysis of our own team-based instructional change effort alone does not present sufficient evidence from which to make strategic decisions, it does illustrate and corroborate prior research that administrators and other change agents could consider when looking to support or initiate these kinds of instructional change efforts in the future.

Our preliminary research also leaves many unanswered questions about our astronomy students and the teaching and learning of undergraduate astronomy majors more broadly. When it comes to student outcomes, while our results seem positive in aggregate, our current data is limited in scope and thus leaves open questions about whether the current form of instruction is helping all students in equitable ways. In particular, we cannot say whether we are achieving (or at least working towards) outcomes consistent with "equity of parity" [Rodriguez et al., 2012], where all students leave the course with comparable knowledge and skills, which to us would be ideal. We did not have sufficient numbers of students (and/or appropriate demographic data) to analyze our data with respect to demographic variables such as race/ethnicity or gender, which would have helped us to start to assess equity in instruction. Similarly, while we suspect that students' prior mathematical preparation and programming experience may influence their performance in the course (and the lab specifically), at present we have no measurement that captures these aspects of students' incoming knowledge and abilities. In the future, we hope to be able to at least look for more markers of *in*equity in the data we collect. Our current data also cannot reveal if and how our efforts might influence students' performance in later classes and their persistence in the major, and improving these aspects of students' experiences is ultimately a goal of our instructional changes. Longitudinal work would be necessary to answer these questions, and would require tracking these students over many semesters and comparing to student data from previous years.

Our work also points to potential footholds and unanswered questions about teaching and researching astronomy majors that could be taken up within the broader astronomy education community. When it comes to research-based assessments, we found initial evidence that the LSCI seems to be valid for astronomy majors through their first year of undergraduate instruction. However, while we suspect that improved conceptual understanding does improve students' problem-solving abilities in astronomy, we currently lack data to support this hypothesis within our course context. Furthermore, we have not experimented with other existing conceptual inventories in astronomy, nor have we attempted to measure other potentially important aspects of student learning within the non-lab portion of the course, such as their attitudes towards participating in astronomy more generally.

When it comes to measuring students' experiences in astronomy labs, we lacked an appropriate instrument and thus modified the E-CLASS survey. However, because the nature of astronomy labs is quite different from physics labs, and because we made modifications to a validated instrument, we suspect that the instrument we used yields some misleading or inaccurate results and is not appropriately suited to measure outcomes that are most consequential within astronomy. Thus, more informative data on students' perceptions of astronomy labs could be obtained if future researchers pursued the development and validation of such an instrument.

Looking beyond astronomy assessments, we note that there is still significant work that could be done to develop instructional models that match the content and skills that one might aim to teach astronomy majors in introductory courses.

While some existing Astro 101 materials such as Lecture-Tutorials and TPS questions were useful and appropriate for teaching astronomy majors, we find ourselves still exploring ways to integrate these existing materials with more complex mathematical tasks that we also included in this course sequence. Moreover, the new and adapted course materials that we did develop revealed aspects of students' thinking that have not been carefully researched or discussed in the literature. For instance, a Lecture-Tutorial-inspired activity that asked students to draw a variety of astronomical objects on a representation of the Milky Way galaxy suggested a breadth of student ideas about how these objects are distributed, whereas prior research only considers students' ideas about the locations of stars relative to the spiral arms. While guiding students to navigate and refine their ideas remains an immediate instructional challenge for our local instructors, these responses also point to potentially shared challenges in structuring learning environments for astronomy majors that could be pursued in a more targeted way. We also suggest that the curricular materials and course structure we developed, both for the main portion of the course and the lab in particular, present useful building blocks and frameworks that others could take up or re-envision to fit within their instruction. Because so little research exists on classroom instruction for undergraduate astronomy majors, we hope that this work presents a modest, new foothold for understanding how we can support a new generation of potential astronomers.

Chapter 7: Conclusions

We began this work motivated by a clear problem in astronomy and physics. Our disciplines, like most STEM fields, are not equally accessible to all people, and commonly used, traditional instructional approaches that heavily rely on lecture perpetuate this inequity at the undergraduate level. While faculty may be more comfortable with traditional instruction because they experienced this as students, they also have significant freedom in how they teach now. We agree with national policy-makers that research-based instructional strategies (RBIS) and education research principles present valuable footholds for promoting equity in the classroom, and that more widespread use of RBIS would likely benefit students. At the same time, we argued that there is much that we still do not know about how to support all students' learning, and that being fully responsive to students will always require instructors to be somewhat flexible. Thus, we argued that education researchers and professional development leaders should both teach faculty what we know now and treat them as partners in instructional change, drawing on their developing expertise and supporting them in making sensible modifications to existing strategies. It is this process of teaching faculty about research-based instruction that we set out to investigate.

Henderson et al. [2011] previously identified four broad approaches to instructional change in undergraduate STEM, which provides an empirical foundation for our work. Two dimensions define the boundaries between these change approaches: whether the change efforts target individuals or groups, and whether they target prescribed or emergent outcomes. Disseminating curriculum and pedagogy—an individually-focused, prescriptive, change approach—is the most common one taken by astronomy and physics education researchers, and many faculty attend professional development (PD) workshops with this primary focus. We know that disseminating highly prescribed RBIS (while discouraging faculty from thinking about how to adapt them) would be inconsistent with treating faculty as partners in the change process, but we argue that some workshop leaders might take more nuanced approaches that could promote more emergent outcomes. This motivated us to study faculty PD workshops in depth, with an eye to these variations in workshop design that could support faculty in making sense of research-based instruction. Additionally, I found an opportunity to pursue a local course transformation project with Derek (one of my co-advisors), which I took up in an attempt to improve the experiences of astronomy majors more directly. In these collaborative efforts, I strove to treat Derek and other instructors as partners during this process, which contributed to our change effort being accurately characterized as "developing shared vision" in this empirical framework. We considered what we learned from enacting this collaborative course transformation effort towards the end of this thesis.

In this final chapter, we will summarize what we learned about faculty instructional change through this research, how we think PD workshops could be re-envisioned in a broad sense, how our work supports this re-envisioning of PD for workshop leaders, and what might comprise promising opportunities for future research that build on our work.

7.1 What have we learned about how we are supporting instructional change?

Prior to our work, very little research had been done to capture what occurs during faculty PD workshops. We have argued that documenting how PD workshops are structured presents a viable starting point for thinking about PD design in a scholarly way. More specifically, we draw from [Sandoval, 2014]'s analytical framework to conceptualize PD design as comprised of embodiment (how workshops are structured), mediating processes (faculty's interactions and engagement within workshops), and outcomes (how faculty change as a result of attending workshops).

We began by looking at embodiment. As we describe in Chapter 4, we created the Real-time Professional Development Observation Tool (R-PDOT) to document the structure of faculty's participation in workshop sessions and the kinds of activities and ideas that faculty are asked to engage with. In creating descriptive codes that capture the broad variations we expect to see within these two aspects of workshop design, we highlighted how all faculty PD workshops are likely comprised of structural elements that could support faculty in improving their instruction to some extent, if assembled and implemented in particular ways.

While developing the R-PDOT, we noticed that many aspects of workshop design that we and other PD researchers highly value because of their potential to support instructors in developing flexible pedagogical expertise were quite rare. In Chapter 4, we showcased R-PDOT data for two sessions of the New Faculty Workshop (NFW) (Sessions A and B) that represent many of the structural features we found to be highly prevalent based on our extensive informal and/or preliminary observations and field notes, and one that contains potentially worthwhile but particularly rare design features (Session C). In particular, we noticed that a majority of NFW sessions (like Sessions A and B) are lecture-heavy, non-interactive, and/or authoritative, leaving little space for faculty to grapple with how pedagogical ideas and strategies from the workshop could be used productively within their local contexts. In looking at the tasks and ideas presented to faculty, we notice that a majority of NFW sessions seem likely to raise faculty's awareness of instructional strategies by describing them and explaining their purpose (like Session A), but seem less likely to support the ambitious PD goals we orient to. Some workshop leaders do create similarity between the workshop contexts and faculty's local contexts by asking faculty to engage in research-based science tasks as pseudo-students. However, it was more unusual to see workshop activities that could help faculty to move from one context to the other, such as analyzing these in situ experiences (which is relatively unusual, and shown in both Sessions B and C) or explicitly planning for future teaching (which was extremely unusual). Moreover, we noticed that open and small group discussions where faculty's pedagogical ideas might be compared, contrasted, and debated are also rare, and we showed only one promising example of how a session might be structured in this way (Session C). Thus, based on this broad picture of how the NFW sessions are typically structured, we suspect that the many common approaches to workshop design are unlikely to promote the ambitious PD goals we orient to.

While R-PDOT data provides a general sense of how sessions are structured, it does not illustrate how productive faculty interactions can be supported or launched, nor does it guarantee that productive faculty interactions occur when promising codes are selected. In Chapter 5, we took a closer look at workshop designs that seem like they could support faculty in developing useful pedagogical expertise. Specifically, we pursued detailed, qualitative video analysis of several instances where faculty experience and reflect on concrete examples of instruction during the NFW. We then made conjectures about how the kinds of faculty interactions we observed (mediating processes) were linked to features of the session design (embodiment), particularly the workshop leader's facilitation moves. Through this analysis, we found evidence that the workshop leader's implementation matters for the nature of faculty's interactions and therefore faculty's opportunities to learn during workshop sessions. We also see evidence that if tasks are well-structured and workshop leaders are attentive to the quality of faculty's interactions and invite faculty to reason about instruction, then faculty seem likely to engage in pedagogical sense-making, e.g., elaborating on how and why various instructional choices might support or constrain students' learning. However, we also see evidence that in these same kinds of well-structured tasks, when workshop leaders do not attend to the quality of instructors' collaboration and/or exclusively focus on adoption of prescribed RBIS, then faculty may act as if they are "doing school," e.g., going through the assigned steps in order to appease the workshop leader, potentially struggling to collaborate with their peers, and not seeking or contributing pedagogical explanations.

We also learned about instructional change by reflecting on our enactment of a local course transformation effort. In Chapter 6, we describe how we redesigned an introductory course sequence for astronomy majors through a team-based effort, and generated new insights into teaching astronomy majors. The successes of this effort support our claims about the value of treating faculty (and other instructors) as partners in instructional change: by treating other instructors as partners, we can accomplish more both in research and in the classroom than we would be able to alone. Team-based change is also still a fairly under-explored area of research, and our experiences corroborate other researchers' emerging ideas about what makes these efforts successful. Moreover, we note that regular, well-structured team meetings seemed to support the successes of this project, as they allowed us to invite input from and distribute substantive responsibilities to many team members. Not coincidentally, there were strong similarities between the activities we focused on during team meetings and focus-of-engagement R-PDOT codes that we consider to be highly valuable, which lends weight to the utility of these codes. While it is unclear if and to what extent these changes will be sustained by future instructors, overall, we claim that we improved the experiences of these particular astronomy students at the University of Maryland and fostered pedagogical expertise in our team members by working with them to identify and create viable instructional solutions.

7.2 Where could we go next?

7.2.1 Our vision for PD workshops

Within these chapters, we have identified many limitations to highly prescriptive, authoritative, and/or non-interactive faculty PD, which complements the previous findings of Henderson et al. [2011] and Henderson et al. [2012]. We do think that increasing faculty's awareness of RBIS is a useful goal, and that disciplinary workshops in particular create valuable opportunities to learn about specific pedagogical strategies and materials that could be well-matched to faculty's instructional needs. However, we also see significant value in teaching faculty to notice their students' ideas and behaviors and adapt their instruction appropriately. We recognize that a single workshop experience alone will not be sufficient for faculty to substantially improve in these ways, but we also think that it is worthwhile to foster faculty's engagement in the kinds of conversations and activities we might want them to continue over long timescales during workshops. This might motivate faculty to pursue these kinds of conversations with other educators in the future and provide them with viable models for how productive instructional collaboration might be structured. We argue that PD workshops should be re-conceptualized as launching points for faculty's long term collaborative learning.

While we have made similar arguments about the purpose of PD at multiple points throughout this thesis, here, we will add that this re-conceptualization of faculty PD workshops is justified given the shifting landscape of higher education.

Unlike in the past, it is often not the case that faculty are isolated from peers who are also trying out RBIS and removed from additional PD opportunities. Many faculty are now aware of RBIS and have experience trying aspects of these strategies in their classrooms, and this will only continue to increase over time. Discipline-based education research is also a growing field, and therefore more departments have STEM education researchers within them who could act as resources for local faculty. Moreover, over the past few years, team-based approaches to instructional change, often at the institutional level, have become increasingly prevalent and are being enabled by the increased national attention to improving classroom instruction in STEM [President's Council of Advisors on Science and Technology, 2012, Weaver et al., 2016]. Thus, it is becoming increasingly plausible that faculty could participate in local change efforts or nationally-distributed communities if they became motivated to do so.

7.2.2 Next steps

In this thesis, we have provided tools and analytical approaches that can support movement towards this vision for PD. In an immediate sense, some of our work can directly support workshop leaders in envisioning and enacting changes to PD design. As we describe in Chapter 4, a central goal of the R-PDOT is to help workshop leaders to reflect on design, and we spent considerable effort developing intuitive visual outputs for this purpose. The aspects of workshop design we chose to highlight with the tool might subtly guide workshop leaders to think about design in new ways that are better aligned with our vision, while workshop leaders who already have goals strongly aligned with ours could use the R-PDOT to assess the consistency of their session design with this vision. Similarly, workshop leaders can use the R-PDOT data to compare and contrast different session designs with others, in ways that are analogous to how we might hope to support faculty's reflective conversations around instruction. Beyond this, we suggest that our analytical framework for thinking about workshop design could be useful for workshop leaders in planning and assessing their sessions. In Chapter 5, we present concrete examples of the kinds of facilitation moves that workshop leaders may (or may not) wish to implement in order to promote faculty's sense-making around instruction and not cue up faculty's school-like ideas and behaviors. Similarly, we provide concrete markers of faculty's engagement in workshop sessions that workshop leaders could attend to, including signs that may indicate that it would be productive for workshop leaders to intervene and redirect faculty's conversations.

Our research also provides some groundwork for future research. Now that the R-PDOT has been developed, future researchers could use this tool to more systematically look across multiple workshop settings and better map out the current landscape of approaches to faculty PD. In illustrating how the R-PDOT could be used, we only considered three session designs, and the variety of additional designs that could be enacted and captured with the R-PDOT leaves much for future researchers to explore. Furthermore, while our qualitative analysis of workshop sessions in Chapter 5 was quite fruitful, it also left many open questions about consequential features of workshop design. The design conjectures we laid out could be refined and made more robust by looking across additional sessions with similar characteristics. There are also alternative session designs (at the NFW and elsewhere) that could also support rich faculty interactions, such as interactions surrounding faculty analyzing and creating student tasks or analyzing students' perspectives together, which could lead to complementary ambitious PD outcomes. It would likely be fruitful to analyze the substance of faculty's talk and engagement within these sessions.

More broadly, our work also could inspire research in team-based change. As we saw in Chapter 5, the ways that faculty's interactions are structured and supported can be highly consequential for the quality and productivity of these interactions. We highlighted some aspects of how instructional collaboration could be structured in Chapter 6, but more formal qualitative research would be needed to understand the affordances and limitations of the different ways that team-based efforts could be supported and facilitated. Moreover, we might imagine that there are some similar affordances and drawbacks to prescriptive versus emergent team-based instructional change efforts, but these have not yet been explored. Given the potential for these team-based efforts to support faculty's long term learning, research in this area would be highly worthwhile.

Ultimately, while we have taken some steps forwards, there is still significant work to be done in both faculty PD research and design. If we are to make the best use of our opportunities to support faculty in improving their teaching and motivate them to engage in future learning opportunities, this kind of work will continue to be necessary and relevant. We hope that researchers and PD leaders will discover new ways to move forward with faculty as we gain insights about how to improve equity in our classroom spaces together.

Appendix A: Extended R-PDOT codebook: Type-of-engagement

Code name	*Code description*	*Example observed behaviors and actions*
Workshop Leader Lecture	Workshop leader lectures while faculty participants listen.	-Workshop leader lecturing on a pre-determined topic. -Workshop leader asking a rhetorical question.
Large Group Closed, Faculty Participant Question	Large group closed discussion. Faculty participant(s) question the workshop leader, and (optionally) the workshop leader answers through lecturing, while all other participants listen.	-Faculty participants challenging the workshop leaders' ideas. -A workshop leader engaging in conversation with a questioner(s), with little or no participant-participant interaction. -A workshop leader addressing a participant's question at length, i.e., the workshop leader's turns of talk are significantly longer than faculty participants' turns of talk.
Large Group Closed, Workshop Leader Question	Large group closed discussion. Workshop leader asks non-rhetorical closed questions, and (optionally) one or more faculty participants respond directly to the workshop leader [Dancy and Henderson, 2007].	-Faculty participants voting in response to a multiple-choice question. -Participants contributing in short phrases or full sentences in response to a workshop leader's question. -Workshop leader and a few faculty participants taking turns speaking, with the discourse focused on revealing an answer the workshop leader is seeking. -A workshop leader revoicing participant responses. -A workshop leader intentionally making an implementation "error" and expecting participants to pause him/her. -A workshop leader reading a question or prompt aloud.
Large Group Open Discuss	Large group open discussion. Workshop leader and faculty participants take turns speaking, with the discourse focused on the ideas of faculty participants [Dancy and Henderson, 2007].	-Faculty participants speaking in complete sentences and contributing new ideas. -A workshop leader asking open-ended questions and facilitating discussion between faculty participants. -Workshop leader(s) participating in discussion in a non-authoritative manner, e.g., sharing their own experiences without privileging them over participants' experiences.

Continued on next page...

Code name	*Code description*	*Example observed behaviors and actions*
Small Group Discuss	Faculty participants discuss with each other in small groups.	-Faculty discussing with their peers during Peer Instruction. -Faculty collaborating on assigned tasks. -Faculty discussing ideas about teaching and learning in response to a prompt. -A workshop leader giving brief instructions while most faculty are still discussing with each other.
Faculty Participant Present	One or more faculty participants present to all others.	-Faculty participants reporting out after working on a task in small groups. -Faculty participants reporting out after working independently. -Workshop leader facilitating presentations, e.g., calling on the next group to speak or revoicing faculty's contributions.
Faculty Participant Independent Work	Faculty participants work independently on a task.	-Faculty participants silently reading a Peer Instruction question. -Faculty participants watching classroom video. -Faculty participants writing independently in response to a prompt or task.

Appendix B: Extended R-PDOT codebook: Focus-of-engagement

Code name	*Code description*	*Example observed behaviors and actions*
Workshop Instructions	Workshop leader instructs participants about what they should or will be doing during the workshop (or participants attempt to clarify these instructions).	-Workshop leader describing the intended purpose of the workshop or workshop session. -Workshop leader telling faculty how to act or what to focus on during an upcoming workshop task. -Faculty recounting or attempting to clarify the instructions for a workshop task.
Education Research Theory and Results	Workshop leader and/or participants emphasize discipline-based education research processes, principles, or findings.	-Describing student misconceptions identified in research. -Showing evidence of improved student outcomes in active learning environments. -Discussing the implications of active learning for diverse student populations. -Describing (workshop leader) or making sense of (faculty participants) education research methods and results. -Describing the characteristics/demographics of students within a particular study. -Describing or discussing research-motivated principles in relation to teaching decisions.

Continued on next page...

Code name	*Code description*	*Example observed behaviors and actions*
Instructional Strategies (IS) Description and Purpose	Workshop leader and/or participants show or describe active learning strategies ranging from current instructional practices to strongly research-based instructional strategies.	-Listing and describing research-based instructional strategies. -Providing instructions about where to find existing materials or questions. -Showing the specific steps that comprise an active learning instructional strategy. -Describing or discussing the purpose of any active learning instructional strategy or components within it. -Describing or discussing the functionality associated with research-based educational technologies (such as PhET sims). -Displaying and briefly discussing research-based questions or tasks. -Describing or discussing possible outcomes of using a particular research-based instructional strategy.
Workshop Leader (WL) Simulating IS	Faculty participants experience a workshop leader's implementation of an instructional strategy, either by acting as mock students, or through observing video, transcript, or case study narrative.	-A workshop leader assuming the role of a science instructor to implement a research-based instructional strategy. -A workshop leader assuming the role of a science instructor to demonstrate common faculty practices. -A workshop leader showing classroom video. -Faculty participants reading a detailed classroom case study. -Faculty participants working through a science task as mock students.
Faculty Participant (FP) Simulating IS (as educator)	A predetermined subset of faculty participants (one or more) try out implementing an instructional strategy while other participants act as mock students.	-Individual faculty participants implementing a research-based instructional strategy with others acting as students. -Teams of faculty participants implementing a research-based instructional strategy with others acting as students. -Faculty participants rehearsing the implementation of a research-based instructional strategy in small groups.

Continued on next page...

Code name	*Code description*	*Example observed behaviors and actions*
Analyzing Simulated IS	Workshop leader and/or participants reflect on (analyze, critique, evaluate, justify) a shared experience of someone simulating an instructional strategy in situ.	-Analyzing how students responded to a given instructor move or reasoned through a science task. -Analyzing the behaviors and actions of an instructor observed during the workshop session. -Reflecting on the quality of faculty participants' engagement in the workshop. -Predicting how students would respond to alternative instructor moves compared to those seen or experienced. -Generating possible next steps an instructor could take following a scenario experienced or considered in the workshop. -Analyzing a hypothetical situation that is closely related to, and generated in relation to, an in situ implementation of an instructional strategy.
WL Pre-Workshop Experiences	Workshop leader and/or participants discuss a workshop leader's past experiences, including instructional goals, practices, values, and local contexts.	-A workshop leader describing their involvement in discipline-based education research and prior studies they have done, using first-person narrative. -A workshop leader describing their teaching practices. -Participants eliciting a workshop leader's past experiences, such as the details of how they teach and the characteristics of their students.
FP Pre-Workshop Experiences	Workshop leader and/or participants reflect on participants' past experiences, including instructional goals, practices, values, and local contexts.	-Faculty participants describing specific teaching strategies they have implemented in their classrooms. -A workshop leader eliciting or stating assumptions about faculty participants' prior teaching and learning experiences. -Faculty participants describing their local teaching resources or constraints.
Student Experiences	Workshop leader and/or participants consider students' knowledge, skills, or affect.	-Hypothesizing about or describing known student explanations of science ideas. -Analyzing real student discourse or student work. -Describing student resistance or buy-in to different instructional approaches and discussing possible causes. -Discussing students' disciplinary knowledge and skills, and considering implications for instruction. -Describing student demographics or characteristics

Continued on next page...

Code name	*Code description*	*Example observed behaviors and actions*
Disciplinary Content Knowledge	Workshop leader and/or participants consider disciplinary ideas.	-Faculty participants engaging in disciplinary tasks using skills and knowledge that students would likely use. -Faculty participants making sense of a task using a broad range of skills and knowledge (potentially including advanced disciplinary ideas). -A workshop leader lecturing or asking questions that require knowledge from the participants' own discipline, e.g., as part of simulating a research-based instructional strategy.
Analyzing and/or Creating Student Tasks	Participants create, modify, or evaluate and/or workshop leaders critique or evaluate specific materials, questions, or tasks for students.	-Writing new questions during the workshop. -Critiquing existing questions or tasks. -Highlighting effective aspects of existing specific research-based materials, questions, or tasks. -Classifying questions using Bloom's taxonomy. -Exploring the pedagogical affordances or constraints of a specific instructional strategy simulation.
Planning for FP Future Teaching	Workshop leader advises and/or faculty participants plan next steps for when participants go back to their home institutions.	-Discussing possible ways to adapt research-based instructional strategies to fit participants' local contexts. -Advising faculty participants about how to approach the process of changing their teaching. -Faculty participants identifying questions or tasks that are particularly relevant for their own instructional goals.

Appendix C: Think-Pair-Share implementation rubric

Implementation Items:

- Did the presenter refrain from reading the question to the students?
- Did the presenter allow time for the students to read and think about the question?
- Did the presenter ask Do you need more time? before going to the first vote?
- Did the presenter get the students to vote simultaneously and anonymously?
- Did the presenter appropriately choose to disclose the distribution of answers from the first vote?
- Did the presenter appropriately direct the students to engage in discourse about their answer choices and explain their reasoning using a prompt that would foster an active discussion?
- Did the presenter use a prompt about the amount of time students would be allowed to collaborate as a way to encourage discussion?
- Did the presenter observe the level and type of student discussions so as to appropriately gauge the amount of time students would need to defend their votes and explain their reasoning?
- Did the presenter provide a prompt about time so students knew their time to discuss would shortly be coming to an end?
- Did the presenter get the students to vote a second time simultaneously and anonymously?
- Did the presenter debrief the final vote results with the students in a pedagogically useful way?

Question Items:

- Did the question serve as a good vehicle to promote a cognitively engaging and conceptually rich discussion amongst the target population?
- Were the answer choices distinct, and representative of likely student conceptual and reasoning difficulties, which a real student might vote for?

Appendix D: Timeline surrounding the ASTR120/121 course transformation

Table D.1: Note that Allison Bostrom, Fatima Abdurrahman, Sarah Scott, Emily Garhart, and Joe DeMartini are all undergraduate astronomy students; Sara Frederick and Holly Sheets are astronomy graduate students. Derek Richardson is the project lead, and Alice Olmstead and Melissa Hayes-Gehrke have advisory roles.

Semester(s)	*Key events*
Fall 2011	Derek teaches ASTR120 Alice TAs for ASTR120 (discussion) Allison and Fatima experience ASTR120 as students
Spring 2012	Derek teaches ASTR121 Holly TAs for ASTR121 (discussion and lab) Allison experiences ASTR121 as a student
Fall 2012	Derek teaches ASTR120 Alice substitute teaches in ASTR120
Spring 2013	Derek and Alice co-teach ASTR121 Holly TAs for ASTR121 (lab) Alice starts to pursue AER Alice briefly tries out TPS in ASTR121 while Derek observes
Summer 2013–Summer 2014	Derek on sabbatical (not teaching) Emily experiences ASTR120 as a student Emily and Fatima experience ASTR121 as students Fatima takes LA pedagogy course and facilitates student work in the lab section a transformed introductory physics for life sciences courses (Fall 2013 and Spring 2014).
Fall 2014	Derek teaches ASTR120 Allison TAs for ASTR120 (discussion) Derek and Allison start to incorporate RBIS into ASTR120 Allison initiates and leads peer tutoring program for undergraduate astronomy majors Alice teaches Astro 101 with LAs (including Sarah and Fatima) Sarah and Fatima take LA pedagogy course Fatima experiences an upper-level astronomy lab as Melissa's student Joe experiences ASTR120 as a student Derek, Alice, and Melissa write TLTC proposal

Continued on next page...

Semester(s)	*Key events*
January 2015	Team obtains 1 year of support from the TLTC Fatima hired onto the team Fatima, Derek, Alice, and Melissa re-envision and begin revising the ASTR121 lab
Spring 2015	Derek teaches ASTR121, continuing to incorporate RBIS (to a limited extent) Allison re-hired as a TA using department funds Sarah hired onto the team Holly informally joins the team to help with lab revisions Regular team meetings begin (Derek, Alice, Melissa, Fatima, Sara, Allison, Holly) Team implements and continues to redesign the ASTR121 lab Joe experiences ASTR121 as a student Derek, Melissa, and Alice participate in TLTC bi-weekly meetings Alice and Fatima attend the 2015 Mid-Atlantic Regional LA Workshop
Summer 2015	ASTR121 lab materials made publicly available Team presents at national AAPT meeting (Derek, Fatima, Alice)
Fall 2015	Derek teaches ASTR120 Sara hired onto the team, leads discussion and attends lecture, audits LA pedagogy course Emily hired as a grader Fatima re-hired for hourly work Weekly team meetings continue (Derek, Alice, Melissa, Sara, Emily, sometimes Fatima, sometimes Holly) Team focuses on making changes to lecture Derek, Melissa, and Alice attend monthly TLTC fellows meetings Derek presents at departmental seminar (BANG!)
Spring 2016	Derek presents at TLTC kick-off meeting Derek teaches ASTR121 Sara retained as a TA Joe recruited and hired to lead ASTR121 lab Formal TLTC support ends Weekly team meetings continue (Derek, Alice, Melissa, Sara, Joe, Holly) Team continue to revise lecture and discussion components of the course Minor improvements made to lab Derek and Alice present at 2016 Mid-Atlantic Regional LA Workshop Derek, Sara, and Joe present at TLTC conference Derek presents at Astronomy faculty meeting

Bibliography

S. L. Adamson, D. Banks, M. Burtch, F. Cox, E. Judson, J. B. Turley, R. Benford, and A. E. Lawson. Reformed undergraduate instruction and its subsequent impact on secondary school teaching practice and student achievement. *Journal of Research in Science Teaching*, 40(10):939–957, dec 2003. ISSN 0022-4308. doi: 10.1002/tea.10117. URL http://doi.wiley.com/10.1002/tea.10117.

A. Amrein-Beardsley and S. E. Osborn Popp. Peer observations among faculty in a college of education: investigating the summative and formative uses of the Reformed Teaching Observation Protocol (RTOP). *Educational Assessment, Evaluation and Accountability*, 24(1):5–24, dec 2011. ISSN 1874-8597. doi: 10.1007/s11092-011-9135-1. URL http://link.springer.com/10.1007/s11092-011-9135-1.

P. Aubusson, J. Griffin, and F. Steele. A Design-Based Self-Study of the Development of Student Reflection in Teacher Education. *Studying Teacher Education*, 6(2):201–216, jul 2010. URL http://www.tandfonline.com/doi/abs/10.1080/17425964.2010.495905.

J. M. Bailey, B. Johnson, E. E. Prather, and T. F. Slater. Development and Validation of the Star Properties Concept Inventory. *International Journal of Science Education*, (November 2015):1–30, 2011. ISSN 1539-1515. doi: 10.3847/AER2006020.

M. M. Bakhtin. Speech genres and other late essays. *Trans. VW McGee. Ed. C. Emerson and M. Holquist. Austin: University of Texas Press*, 1986.

D. Ball and D. Cohen. Developing practice, developing practitioners: Toward a practice-based theory of professional education. In *Teaching as the learning profession: Handbook of policy and practice*, pages 3–32. 1999. ISBN 078794341X (alk. paper). doi: 10.1037/0022-3514.90.4.644.

A. Bandura. *Social foundations of thought and action: A social cognitive theory.* Prentice-Hall, Inc, 1986.

E. M. Bardar, E. E. Prather, K. Brecher, and T. F. Slater. The Need for a Light and Spectroscopy Concept Inventory for Assessing Innovations in Introductory

Astronomy Survey Courses. *Astronomy Education Review*, 4(2):20, 2005. ISSN 1539-1515. doi: 10.3847/AER2005018.

B. Barron. Achieving coordination in collaborative problem-solving groups. *The Journal of the Learning Sciences*, 9(4):403–436, 2000.

B. Barron. When smart groups fail. *The Journal of the Learning Sciences*, 12(3): 307–359, 2003.

A. L. Beach and M. D. Cox. The impact of faculty learning communities on teaching and learning. *Learning Communities Journal*, 1(1):7–27, 2009.

S. L. Beilock, R. J. Rydell, and A. R. McConnell. Stereotype threat and working memory: mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, 136(2):256, 2007.

J. O. Bennett, M. Donahue, N. Schneider, and M. Voit. *The cosmic perspective*. Pearson Addison-Wesley, 2016.

M. Borrego, J. E. Froyd, and T. S. Hall. Diffusion of Engineering Education Innovations: A Survey of Awareness and Adoption Rates in U.S. Engineering Departments. *Journal of Engineering Education*, 99(3):185–207, jul 2010. ISSN 10694730. doi: 10.1002/j.2168-9830.2010.tb01056.x. URL http://doi.wiley.com/10.1002/j.2168-9830.2010.tb01056.x.

J. Bouwma-Gearhart, K. H. Perry, and J. B. Presley. Improving Postsecondary STEM Education: Strategies for Successful Interdisciplinary Collaborations and Brokering Engagement With Education Research and Theory. *Journal of College Science Teaching*, 44(1):40–47, 2014. ISSN 0047231X.

J. D. Bransford, A. L. Brown, and R. R. Cocking. *How People Learn: Brain, Mind, Experience, and School*, volume Expanded E. 2000. ISBN 0309070368. doi: 10.1016/0885-2014(91)90049-J. URL http://www.nap.edu/openbook.php?isbn=0309070368.

E. Brewe, V. Sawtelle, L. H. Kramer, G. E. O'Brien, I. Rodriguez, and P. Pamelá. Toward equity through participation in Modeling Instruction in introductory university physics. *Physical Review Special Topics-Physics Education Research*, 6(1): 10106, 2010.

K. Brodie. Working with learners' mathematical thinking: Towards a language of description for changing pedagogy. *Teaching and Teacher Education*, 27(1): 174–186, 2011.

D. A. Budd, K. J. v. d. H. Kraft, D. A. McConnell, and T. Vislova. Characterizing Teaching in Introductory Geology Courses: Measuring Classroom Practices. *Journal of Geological Education*, 61:461–475, nov 2013. URL http://nagt-jge.org/doi/abs/10.5408/12-381.1.

H. B. Carlone. The cultural production of science in reform-based physics: Girls' access, participation, and resistance. *Journal of Research in Science Teaching*, 41(4):392–414, apr 2004. ISSN 0022-4308. doi: 10.1002/tea.20006. URL http://doi.wiley.com/10.1002/tea.20006.

C. B. Cazden. The language of teaching and learning. *The language of teaching and learning*, 2001.

A. Chappell. Using Teaching Observations and Reflective Practice to Challenge Conventions and Conceptions of Teaching in Geography. *Journal of Geography in Higher Education*, 31(2):257–268, apr 2007. ISSN 0309-8265. doi: 10.1080/03098260601063651. URL http://www.tandfonline.com/doi/abs/10.1080/03098260601063651{#}.Vqe0D5MrKRs.

S. V. Chasteen. Getting the word out: Effective communication of our work in PER, 2011.

S. V. Chasteen, K. K. Perkins, P. D. Beale, S. J. Pollock, and C. E. Wieman. A thoughtful approach to instruction: Course transformation for the rest of us. *Journal of College Science Teaching*, 40(4):24–30, 2011.

S. V. Chasteen, B. Wilcox, M. D. Caballero, K. K. Perkins, S. J. Pollock, and C. E. Wieman. Educational transformation in upper-division physics: The Science Education Initiative model, outcomes, and lessons learned. *Physical Review Special Topics-Physics Education Research*, 11(2):20110, 2015.

N. V. N. Chism. *Peer review of teaching: a sourcebook*. Anker Publishing Co., Bolton, MA, 2nd edition, 2007.

W. Christian and M. Belloni. *Physlet Physics*. Prentice-Hall, Inc., 2003.

P. Cobb, J. Confrey, R. Lehrer, L. Schauble, and Others. Design experiments in educational research. *Educational Researcher*, 32(1):9–13, 2003.

J. E. Coffey, D. Hammer, D. M. Levin, and T. Grant. The missing disciplinary substance of formative assessment. *Journal of Research in Science Teaching*, 48 (10):1109–1136, dec 2011. ISSN 00224308. doi: 10.1002/tea.20440. URL http://doi.wiley.com/10.1002/tea.20440.

D. K. Cohen, S. W. Raudenbush, and D. L. Ball. Resources, Instruction, and Research. *Educational Evaluation and Policy Analysis*, 25(2):119–142, jan 2003. ISSN 0162-3737. doi: 10.3102/01623737025002119. URL http://epa.sagepub.com/content/25/2/119.short.

J. Cohen. Statistical power analysis for the social sciences. 1988.

M. Cole. *Cultural psychology: A once and future discipline*. Harvard University Press, 1998.

L. D. Conlin. Building shared understandings in introductory physics tutorials through risk, repair, conflict & comedy. 2012.

J. C. Corbo, D. L. Reinholz, M. H. Dancy, S. Deetz, and N. Finkelstein. A Framework for Transforming Departmental Culture to Support Educational Innovation. *Physical Review - Physics Education Research*, 12(1):010113, 2016. URL http://arxiv.org/abs/1412.3034.

Council of Scientific Society Presidents. The Role of Scientific Societies in STEM Faculty Workshops. Technical report, Washington, DC, 2012.

G. Dall'Alba. Foreshadowing conceptions of teaching. *Research and development in higher education*, 13:293–297, 1991.

M. Dancy and C. Henderson. Framework for articulating instructional practices and conceptions. *Physical Review Special Topics - Physics Education Research*, 3(1): 010103, may 2007. ISSN 1554-9178. doi: 10.1103/PhysRevSTPER.3.010103. URL http://journals.aps.org/prper/abstract/10.1103/PhysRevSTPER.3.010103.

M. Dancy and C. Henderson. Pedagogical Practices of Physics Faculty in the USA. In *Physics Education Research Conference 2009*, pages 121–124, Ann Arbor, MI, 2009. PER Conference.

M. Dancy and C. Henderson. Pedagogical practices and instructional change of physics faculty. *American Journal of Physics*, 78(10):1056, sep 2010. ISSN 00029505. doi: 10.1119/1.3446763. URL http://scitation.aip.org/content/aapt/journal/ajp/78/10/10.1119/1.3446763.

M. Dancy, C. Henderson, and C. Turpen. How faculty learn about and implement research-based instructional strategies: The case of Peer Instruction. *Physical Review - Physics Education Research*, 12(1):10110, feb 2016. doi: 10.1103/PhysRevPhysEducRes.12.010110. URL http://link.aps.org/doi/10.1103/PhysRevPhysEducRes.12.010110.

M. H. Dancy, C. Turpen, and C. Henderson. Why do faculty try research-based instructional strategies? In *Physics Education Research Conference 2010*, pages 117–120, Portland, OR, 2010. PER Conference.

H. Daniels. *Vygotsky and research*. Routledge, 2008.

L. Darling-Hammond, R. C. Wei, A. Andree, N. Richardson, and S. Orphanos. Professional learning in the learning profession. *Washington, DC: National Staff Development Council*, 2009.

L. M. Desimone, A. C. Porter, M. S. Garet, K. S. Yoon, and B. F. Birman. Effects of Professional Development on Teachers' Instruction: Results from a Three-year Longitudinal Study. *Educational Evaluation and Policy Analysis*, 24(2):

81–112, jan 2002. ISSN 0162-3737. doi: 10.3102/01623737024002081. URL http://epa.sagepub.com/content/24/2/81.short.

S. Deustua, J. N.-S. RIT, and T. Foster. In Support of Astronomy Education Research. *astro2010: The Astronomy and Astrophysics Decadal Survey*, 2010:9P, 2009.

A. A. Disessa and B. L. Sherin. What changes in conceptual change? *International Journal of Science Education*, 20(10):1155–1191, 1998.

E. Duckworth. *The having of wonderful ideas and other essays on teaching and learning*. Teachers College Press, 1996.

D. Ebert-May, T. L. Derting, J. Hodder, J. L. Momsen, T. M. Long, and S. E. Jardeleza. What We Say Is Not What We Do: Effective Evaluation of Faculty Professional Development Programs. *BioScience*, 61(7):550–558, jul 2011. ISSN 00063568. doi: 10.1525/bio.2011.61.7.9. URL http://bioscience.oxfordjournals.org/content/61/7/550.full.

R. A. Engle and F. R. Conant. Guiding Principles for Fostering Productive Disciplinary Engagement: Explaining an Emergent Argument in a Community of Learners Classroom. *Cognition and Instruction*, 20(4):399–483, jun 2010. ISSN 0737-0008. doi: 10.1207/S1532690XCI2004{_}1. URL http://www.tandfonline.com/doi/abs/10.1207/S1532690XCI2004{_}1.

I. Esmonde. Ideas and identities: Supporting equity in cooperative mathematics learning. *Review of Educational Research*, 79(2):1008–1043, 2009.

K. B. Follette, D. W. McCarthy, E. Dokter, S. Buxner, and E. Prather. The Quantitative Reasoning for College Science (QuaRCS) Assessment, 1: Development and Validation. *Numeracy*, 8(2), 2015.

K. Foote, A. Knaub, C. Henderson, M. Dancy, and R. J. Beichner. Enabling and challenging factors in institutional reform: The case of SCALE-UP. *Physical Review - Physics Education Research*, 12(1):10103, feb 2016. doi: 10.1103/PhysRevPhysEducRes.12.010103. URL http://link.aps.org/doi/10.1103/PhysRevPhysEducRes.12.010103.

S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, 111(23):8410–5, jun 2014. ISSN 1091-6490. doi: 10.1073/pnas.1319030111. URL http://www.pnas.org/content/111/23/8410.short.

M. R. Gallos, E. van den Berg, and D. F. Treagust. The effect of integrated course and faculty development: Experiences of a university chemistry department in the Philippines. *International Journal of Science Education*, 27(8):985–1006, jan

2005. ISSN 0950-0693. doi: 10.1080/09500690500038447. URL http://www.tandfonline.com/doi/abs/10.1080/09500690500038447{#}.Vq6XkzYrKRs.

M. S. Garet, A. C. Porter, L. Desimone, B. F. Birman, and K. S. Yoon. What Makes Professional Development Effective? Results From a National Sample of Teachers. *American Educational Research Journal*, 38(4):915–945, jan 2001. ISSN 0002-8312. doi: 10.3102/00028312038004915. URL http://aer.sagepub.com/content/38/4/915.short.

R. M. Goertzen, R. E. Scherr, and A. Elby. Respecting tutorial instructors' beliefs and experiences: A case study of a physics teaching assistant. *Physical Review Special Topics-Physics . . .* , 6(2):020125, 2010.

C. Goodwin. Professional Vision. *American Anthropologist*, 96(3):606–633, 1994. ISSN 0002-7294. doi: 10.1525/aa.1994.96.3.02a00100. URL http://doi.wiley.com/10.1525/aa.1994.96.3.02a00100.

M. H. Goodwin. *He-said-she-said: Talk as Social Organization Among Black Children*. Indiana University Press, 1990. ISBN 0253206189. URL https://books.google.com/books?hl=en{&}lr={&}id=ESDkGWmjNiIC{&}pgis=1.

M. J. Graham, J. Frederick, A. Byars-Winston, A.-B. Hunter, and J. Handelsman. Increasing persistence of college students in STEM. *Science*, 341(6153):1455–1456, 2013.

C. Guarino and B. Stacy. Review of ???Gathering Feedback for Teaching.??? *Boulder, CO: National Education Policy Center.*, 28, 2012.

K. Gutiérrez, D. Berlin, K. Crosland, and A. Razfar. Social organization of learning classroom observation protocol. *Los Angeles: Center for the Study of Urban Literacies, Graduate School of Education and Information Studies, University of California, Los Angeles*, 1999.

R. R. Hake. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics*, 66(1):64–74, 1998.

J. Halley, A. Eshleman, and R. M. Vijaya. *Seeing white: An introduction to white privilege and race*. Rowman & Littlefield Publishers, 2011.

D. Hammer. Student resources for learning introductory physics. *American Journal of Physics*, 68(S1):S52—-S59, 2000.

D. Hammer, A. Elby, R. E. Scherr, and E. F. Redish. Resources, framing, and transfer. *Transfer of learning from a modern multidisciplinary perspective*, pages 89–120, 2005.

D. Hammer, F. Goldberg, and S. Fargason. Responsive teaching and the beginnings of energy in a third grade classroom. *Review of science, mathematics and ICT education*, 6(1):51–72, 2012.

L. Hammersley-Fletcher and P. Orsmond. Evaluating our peers: is peer observation a meaningful process? *Studies in Higher Education*, 29(4):489–503, 2004.

L. Hammersley-Fletcher and P. Orsmond. Reflecting on reflective practices within peer observation. *Studies in Higher Education*, 30(2):213–224, apr 2005. ISSN 0307-5079. doi: 10.1080/03075070500043358. URL http://www.tandfonline.com/doi/abs/10.1080/03075070500043358.

S. E. Hampton and R. A. Reiser. Effects of a Theory-Based Feedback and Consultation Process on Instruction and Learning in College Classrooms. *Research in Higher Education*, 45(5):497–527, aug 2004. ISSN 0361-0365. doi: 10.1023/B:RIHE.0000032326.00426.d5. URL http://link.springer.com/10.1023/B:RIHE.0000032326.00426.d5.

D. B. Harlow. Structures and improvisation for inquiry-based science instruction: A teacher's adaptation of a model of magnetism activity. *Science education*, 94 (1):142–163, 2010.

D. B. Harlow, J. A. Bianchini, L. H. Swanson, and H. A. Dwyer. Potential teachers' appropriate and inappropriate application of pedagogical resources in a model-based physics course: "A knowledge in pieces" perspective on teacher learning. *Journal of Research in Science Teaching*, 50(9):1098–1126, nov 2013. ISSN 00224308. doi: 10.1002/tea.21108. URL http://doi.wiley.com/10.1002/tea.21108.

N. Hativa. The department-wide approach to improving faculty instruction in higher education: A qualitative evaluation. *Research in Higher Education*, 36(4):377–413, aug 1995. ISSN 0361-0365. doi: 10.1007/BF02207904. URL http://link.springer.com/10.1007/BF02207904.

W. D. Hawley and L. Valli. Learner-centered professional development. *Phi Delta Kappa Center for Evaluation, Development, and Research*, 27:7–10, 2000.

Z. Hazari, R. H. Tai, and P. M. Sadler. Gender differences in introductory university physics performance: The influence of high school physics preparation and affective factors. *Science Education*, 91(6):847–876, 2007. ISSN 00368326. doi: 10.1002/sce.20223.

P. Heller and M. Hollabaugh. Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups. *American Journal of Physics*, 60(7):637–644, 1992.

P. Heller, R. Keith, and S. Anderson. Teaching problem solving through cooperative grouping (Part 1): Group versus individual problem solving. *MAA NOTES*, pages 159–172, 1997.

C. Henderson. Promoting instructional change in new faculty: An evaluation of the physics and astronomy new faculty workshop. *American Journal of Physics*, 76 (2):179–187, feb 2008. ISSN 00029505. doi: 10.1119/1.2820393. URL http://scitation.aip.org/content/aapt/journal/ajp/76/2/10.1119/1.2820393.

C. Henderson and M. H. Dancy. Physics faculty and educational researchers: Divergent expectations as barriers to the diffusion of innovations. *American Journal of Physics*, 76(1):79, jan 2008. ISSN 00029505. doi: 10.1119/1.2800352. URL http://scitation.aip.org/content/aapt/journal/ajp/76/1/10.1119/1.2800352.

C. Henderson and M. H. Dancy. Impact of physics education research on the teaching of introductory quantitative physics in the United States. *Physical Review Special Topics - Physics Education Research*, 5(2):020107, dec 2009. ISSN 1554-9178. doi: 10.1103/PhysRevSTPER.5.020107. URL http://journals.aps.org/prper/abstract/10.1103/PhysRevSTPER.5.020107.

C. Henderson, A. Beach, and N. Finkelstein. Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of Research in Science Teaching*, 48(8):952–984, oct 2011. ISSN 00224308. doi: 10.1002/tea.20439. URL http://doi.wiley.com/10.1002/tea.20439.

C. Henderson, M. Dancy, and M. Niewiadomska-Bugaj. Use of research-based instructional strategies in introductory physics: Where do faculty leave the innovation-decision process? *Physical Review Special Topics - Physics Education Research*, 8(2):020104, jul 2012. ISSN 1554-9178. doi: 10.1103/PhysRevSTPER.8.020104. URL http://journals.aps.org/prper/abstract/10.1103/PhysRevSTPER.8.020104.

C. Henderson, C. Turpen, M. Dancy, and T. Chapman. Assessment of teaching effectiveness : Lack of alignment between instructors , institutions , and research recommendations. *Physical Review Special Topics - Physics Education Research*, 10(1):010106, 2014. doi: 10.1103/PhysRevSTPER.10.010106.

C. Hill, C. Corbett, and A. St. Rose. Why So Few? Women in Science, Technology, Engineering, and Mathematics. *American Association of University Women*, nov 2009. URL http://eric.ed.gov/?id=ED509653.

M. T. Hora. Organizational Factors and Instructional Decision-Making: A Cognitive Perspective. *The Review of Higher Education*, 35(2):207–235, 2012. ISSN 1090-7009. doi: 10.1353/rhe.2012.0001. URL https://muse.jhu.edu/journals/review{_}of{_}higher{_}education/v035/35.2.hora.html.

M. T. Hora and J. Ferrare. A review of classroom observation techniques in postsecondary settings (WCER Working Paper 2013-1). 2013. URL http://www.wcer.wisc.edu/publications/workingPapers/papers.php.

M. T. Hora, A. Oleson, and J. J. Ferrare. Teaching Dimensions Observation Protocol (TDOP) User's Manual. *Madison: Wisconsin Center for Education Research*, 2013.

I. S. Horn. Teaching replays, teaching rehearsals, and re-visions of practice: Learning from colleagues in a mathematics teacher community. *The Teachers College Record*, 112(1):225–259, 2010.

I. S. Horn and B. D. Kane. Opportunities for Professional Learning in Mathematics Teacher Workgroup Conversations: Relationships to Instructional Expertise. *Journal of the Learning Sciences*, 24(3):373–418, apr 2015. ISSN 1050-8406. doi: 10.1080/10508406.2015.1034865. URL http://www.tandfonline.com/doi/abs/10.1080/10508406.2015.1034865.

I. S. Horn and J. W. Little. Attending to problems of practice: Routines and resources for professional learning in teachers' workplace interactions. *American Educational Research Journal*, 47(1):181–217, 2010.

D. W. Hudgins, E. E. Prather, D. J. Grayson, and D. P. Smits. Effectiveness of Collaborative Ranking Tasks on Student Understanding of Key Astronomy Concepts. *Astronomy Education Review*, 5(1):1, 2006. doi: 10.3847/AER2006001.

B. Hufnagel. Development of the Astronomy Diagnostic Test. *Astronomy Education Review*, 1(1):47, 2001. ISSN 1539-1515. doi: 10.3847/AER2001004.

P. Hutchings and L. S. Shulman. The scholarship of teaching: New elaborations, new developments. *Change: The Magazine of Higher Learning*, 31(5):10–15, 1999.

E. Hutchins. *Cognition in the Wild*. MIT press, 1995.

M. P. Jimenez-Aleixandre, A. B. Rodriguez, and R. A. Duschl. "Doing the lesson" or "doing science": Argument in high school genetics. *Science Education*, 84(6): 757–792, 2000.

A. C. Johnson. Unintended consequences: How science professors discourage women of color. *Science Education*, 91(5):805–821, 2007.

E. Judson and A. E. Lawson. What is the role of constructivist teachers within faculty communication networks? *Journal of Research in Science Teaching*, 44 (3):490–505, mar 2007. ISSN 00224308. doi: 10.1002/tea.20117. URL http://doi.wiley.com/10.1002/tea.20117.

A. S. Jurow. Shifting Engagements in Figured Worlds: Middle School Mathematics Students' Participation in an Architectural Design Project. *Journal of the Learning Sciences*, 14(1):35–67, jan 2005. ISSN 1050-8406. doi: 10.1207/s15327809jls1401{_}3. URL http://www.tandfonline.com/doi/abs/10.1207/s15327809jls1401{_}3.

T. J. Kane and D. O. Staiger. Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains. Research Paper. MET Project. *Bill & Melinda Gates Foundation*, dec 2011. URL http://eric.ed.gov/?id=ED540960.

J. M. Keller. *Part I. Development of a concept inventory addressing students' beliefs and reasoning difficulties regarding the greenhouse effect, Part II. Distribution of chlorine measured by the Mars Odyssey Gamma Ray Spectrometer*. 2006.

D. Kember. A reconceptualisation of the research into university academics' conceptions of teaching. *Learning and Instruction*, 7(3):255–275, sep 1997. ISSN 09594752. doi: 10.1016/S0959-4752(96)00028-X. URL http://www.sciencedirect.com/science/article/pii/S095947529600028X.

A. Kezar. Understanding and facilitating organizational change in the 21st century. *ASHE-ERIC higher education report*, 28(4):147, 2001.

E. Knuth and D. Peressini. Unpacking the nature of discourse in mathematics classrooms. *Mathematics Teaching in the Middle School*, 6(5):320, 2001.

L. Kost, S. Pollock, and N. Finkelstein. Characterizing the gender gap in introductory physics. *Physical Review Special Topics - Physics Education Research*, 5(1): 1–14, 2009. ISSN 1554-9178. doi: 10.1103/PhysRevSTPER.5.010101.

J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.

E. S. Lane and S. E. Harris. A new tool for measuring student behavioral engagement in large university classes. *Journal of College Science Teaching*, 44(6):83–91, 2015.

J. Lave. Teaching, as learning, in practice. *Mind, culture, and activity*, 3(3):149–164, 1996.

P. Laws. Workshop Physics: Learning introductory physics by doing it. *Change: The Magazine of Higher Learning*, 23(4):20–27, 1991.

J. L. Lemke. Talking Science: Language, Learning, and Values. nov 1989. URL http://eric.ed.gov/?id=ED362379.

R. S. Lindell and J. P. Olsen. Developing the lunar phases concept inventory. In *Proceedings of the 2002 Physics Education Research Conference*. New York: PERC Publishing, 2002.

D. C. Lortie and D. Clement. *Schoolteacher: A sociological study*. University of Chicago Press, Chicago, 1975.

Y. M. Lotman. Text within a text. *Soviet psychology*, 26(3):32–51, 1988.

S. Loucks-Horsley, K. E. Stiles, S. Mundry, N. Love, and P. W. Hewson. *Designing Professional Development for Teachers of Science and Mathematics*, volume 24. SAGE Publications, 2009. ISBN 1452208298. URL https://books.google.com/books?hl=en{&}lr={&}id=GSd2AwAAQBAJ{&}pgis=1.

T. J. Lund and M. Stains. The importance of context: an exploration of factors influencing the adoption of student-centered teaching among chemistry, biology, and physics faculty. *International Journal of STEM Education*, 2(1): 13, aug 2015. ISSN 2196-7822. doi: 10.1186/s40594-015-0026-8. URL http://www.stemeducationjournal.com/content/2/1/13.

T. J. Lund, M. Pilarz, J. B. Velasco, D. Chakraverty, K. Rosploch, M. Undersander, and M. Stains. The best of both worlds: Building on the COPUS and RTOP observation protocols to easily and reliably measure various levels of reformed instructional practice. *CBE-Life Sciences Education*, 14(2), jan 2015. ISSN 1931-7913. doi: 10.1187/cbe.14-10-0168. URL http://www.lifescied.org/content/14/2/ar18.short.

R. H. Macdonald, C. A. Manduca, D. W. Mogk, and B. J. Tewksbury. Teaching methods in undergraduate geoscience courses: Results of the 2004 On the Cutting Edge survey of US faculty. *Journal of Geoscience Education*, 53(3):237, 2005.

D. MacIsaac. Reforming Physics Instruction Via RTOP. *The Physics Teacher*, 40(8):479, dec 2002. ISSN 0031921X. doi: 10.1119/1.1526620. URL http://scitation.aip.org/content/aapt/journal/tpt/40/8/10.1119/1.1526620.

D. MacIsaac, D. Sawada, and K. Falconer. Using the Reformed Teaching Observation Protocol (RTOP) as a Catalyst for Self-Reflective Change in Secondary Science Teaching. Technical report, Seattle, WA, apr 2001. URL http://eric.ed.gov/?id=ED452070.

G. MacLachlan and I. Reid. Framing and interpretation. 1994.

A. Madsen, S. B. McKagan, and E. C. Sayre. How physics instruction impacts students' beliefs about learning physics. *Physical Review Special Topics - Physics Education Research*, 11(1):010115, 2015. ISSN 1554-9178. doi: 10.1103/PhysRevSTPER.11.010115. URL http://arxiv.org/abs/1403.6522.

C. A. Manogue, P. J. Siemens, J. Tate, K. Browne, M. L. Niess, and A. J. Wolfer. Paradigms in Physics: A new upper-division curriculum. *American Journal of Physics*, 69(9):978–990, 2001.

J. Margolis and A. Fisher. *Unlocking the clubhouse: Women in computing*. MIT press, 2003.

L. Markauskaite and P. Goodyear. Tapping into the mental resources of teachers' working knowledge: Insights into the generative power of intuitive pedagogy. *Learning, Culture and Social Interaction*, 3(4):237–251, dec 2014. ISSN

22106561. doi: 10.1016/j.lcsi.2014.01.001. URL http://www.sciencedirect.com/science/article/pii/S2210656114000026.

J. C. Marshall, J. Smart, and R. M. Horton. The design and validation of EQUIP: An instrument to assess inquiry-based instruction. *International Journal of Science and Mathematics Education*, 8(2):299–321, 2010.

E. Mazur. *Peer Instruction: A User's Manual*. Prentice-Hall, Inc., Upper Saddle River, NJ, 1997.

L. C. McDermott and P. S. Shaffer. *Tutorials in introductory physics*. Prentice Hall, 1998.

K. L. McNeill, D. J. Lizotte, J. Krajcik, and R. W. Marx. Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences*, 15(2):153–191, 2006.

J. McShannon, P. Hynes, N. Nirmalakhandan, G. Venkataramana, C. Ricketts, A. Ulery, and R. Steiner. Gaining Retention and Achievement for Students Program: A Faculty Development Program. *Journal of Professional Issues in Engineering Education and Practice*, 132(3):204–208, jul 2006. ISSN 1052-3928. doi: 10.1061/(ASCE)1052-3928(2006)132:3(204). URL http://ascelibrary.org/doi/abs/10.1061/(ASCE)1052-3928(2006)132:3(204).

H. Mehan. *Learning lessons*. Harvard University Press Cambridge, MA, 1979.

L. Merner. African American Participation among Bachelors in the Physical Sciences and Engineering. Technical report, AIP Statistical Research Center, College Park, MD, 2015. URL https://www.aip.org/sites/default/files/statistics/minorities/africanamer-pse-13.1.pdf.

P. D. Morrell and A. C. Schepige. *Self-Studies of Science Teacher Education Practices*, volume 12 of *Self-Study of Teaching and Teacher Education Practices*. Springer Netherlands, Dordrecht, 2012. ISBN 978-94-007-3903-1. doi: 10.1007/978-94-007-3904-8. URL http://www.springerlink.com/index/10.1007/978-94-007-3904-8.

P. D. Morrell, C. Wainwright, and L. Flick. Reform Teaching Strategies Used by Student Teachers. *School Science and Mathematics*, 104(5):199–213, may 2004. ISSN 00366803. doi: 10.1111/j.1949-8594.2004.tb18243.x. URL http://doi.wiley.com/10.1111/j.1949-8594.2004.tb18243.x.

P. J. Mulvey and S. Nicholson. Physics Bachelor's Degrees. Technical report, AIP Statistical Research Center, College Park, MD, 2015. URL https://www.aip.org/sites/default/files/statistics/undergrad/bachdegrees-p-14.pdf.

National Research Council. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. National Academies Press,

2012. ISBN 0309217423. URL https://books.google.com/books?hl=en{&}lr={&}id=b2L5VShktGIC{&}pgis=1.

National Science and Technology Council Committee on STEM Education. The federal science, technology, engineering, and mathematics (STEM) 5-year strategic plan. Technical report, 2013. URL https://www.whitehouse.gov/sites/default/files/microsites/ostp/stem{_}stratplan{_}2013.pdf.

G. M. Novak, E. T. Patterson, A. D. Gavrin, and W. Christian. Just-In-Time Teaching Blending Active Learning with Web Technology. 1999.

E. Ochs. Transcription as theory. *Developmental pragmatics*, 10(1):43–72, 1979.

M. C. O'Connor and S. Michaels. Aligning Academic Task and Participation Status through Revoicing: Analysis of a Classroom Discourse Strategy. *Anthropology and Education Quarterly*, 24(4):318–335, dec 1993. ISSN 0161-7761. doi: 10.1525/aeq.1993.24.4.04x0063k. URL http://doi.wiley.com/10.1525/aeq.1993.24.4.04x0063k.

A. Oleson and M. T. Hora. Teaching the way they were taught? Revisiting the sources of teaching knowledge and the role of prior experience in shaping faculty teaching practices. *Higher Education*, 68(1):29–45, oct 2013. ISSN 0018-1560. doi: 10.1007/s10734-013-9678-9. URL http://link.springer.com/10.1007/s10734-013-9678-9.

A. Olmstead and C. Turpen. "I got in trouble": A case study of faculty "doing school" during professional development. In *Physics Education Research Conference 2015*, PER Conference, pages 243–246, College Park, MD, jul 2015.

R. Olson. Don???t be such a scientist. *Island Pr*, 2009.

A. N. Parks and M. Schmeichel. Obstacles to addressing race and ethnicity in the mathematics education literature. *Journal for Research in Mathematics Education*, 43(3):238–252, 2012.

K. Perkin and M. Gratny. Who Becomes a Physics Major? A Long-term Longitudinal Study Examining the Roles of Pre-college Beliefs about Physics and Learning Physics, Interest, and Academic Achievement. In *Physics Education Research Conference 2010*, pages 254–256, Portland, OR, 2010. PER Conference.

K. Perkins, W. Adams, M. Dubson, N. Finkelstein, S. Reid, C. Wieman, and R. LeMaster. PhET: Interactive simulations for teaching and learning physics. *The Physics Teacher*, 44(1):18–23, 2006.

S. Piccinin, C. Cristi, and M. McCoy. The impact of individual consultation on student ratings of teaching. *International Journal for Academic Development*, 4(2):75–88, jul 2006. ISSN 1360-144X. doi: 10.1080/1360144990040202. URL http://www.tandfonline.com/doi/abs/10.1080/1360144990040202{#}.Vq6YpDYrKRs.

D. C. Pope. *Doing school: How we are creating a generation of stressed out, materialistic, and miseducated students*. Yale University Press, 2001.

E. E. Prather and G. Brissenden. Development and Application of a Situated Apprenticeship Approach to Professional Development of Astronomy Instructors. *Astronomy Education Review*, 7(2):1–17, 2008. ISSN 15391515. doi: 10.3847/AER2008016.

E. E. Prather, T. F. Slater, J. P. Adams, J. M. Bailey, L. V. Jones, and J. A. Dostal. Research on a Lecture-Tutorial Approach to Teaching Introductory Astronomy for Non-Science Majors, 2004. ISSN 15391515.

E. E. Prather, A. L. Rudolph, G. Brissenden, and W. M. Schlingman. A national study assessing the teaching and learning of introductory astronomy. Part I. The effect of interactive instruction. *American Journal of Physics*, 77(4):320, 2009. ISSN 00029505. doi: 10.1119/1.3065023.

President's Council of Advisors on Science and Technology. Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics. Report to the President. *Executive Office of the President*, 2012.

E. F. Redish. Implications of cognitive studies for teaching physics, 1994. ISSN 00029505.

E. F. Redish. *Teaching Physics with the Physics Suite*. Wiley, 2003.

E. F. Redish, C. Bauer, K. L. Carleton, T. J. Cooke, M. Cooper, C. H. Crouch, B. W. Dreyfus, B. D. Geller, J. Giannini, J. S. Gouvea, and Others. NEXUS/Physics: An interdisciplinary repurposing of physics for biologists. *American Journal of Physics*, 82(5):368–377, 2014.

J. Richards. Exploring what stabilizes teachers' attention and responsiveness to the substance of students' scientific thinking in the classroom. 2013.

A. D. Robertson. *Responsive Teaching in Science and Mathematics*. Routledge, 2015.

I. Rodriguez, E. Brewe, V. Sawtelle, and L. H. Kramer. Impact of equity models and statistical measures on interpretations of educational reform. *Physical Review Special Topics-Physics Education Research*, 8(2):20103, 2012.

B. Rogoff. *Apprenticeship in thinking: Cognitive development in social context.* Oxford University Press, 1990.

A. L. Rudolph, E. E. Prather, G. Brissenden, D. Consiglio, and V. Gonzaga. A National Study Assessing the Teaching and Learning of Introductory Astronomy Part II: The Connection between Student Demographics and Learning. *Astronomy Education Review*, 9(1):010107, 2010. ISSN 15391515. doi: 10.3847/AER0009068.

A. Rundquist, J. Corbo, S. Chasteen, M. Martinuk, C. Henderson, and M. Dancy. Faculty Online Learning Communities to support physics teaching. In *Physics Education Research Conference 2015*, pages 279–282, College Park, MD, 2015.

M. S. Sabella. Implementing Tutorials in Introductory Physics at an Inner-City University in Chicago. In *Proceedings of the 2002 Physics Education Research Conference, edited by S. Franklin, K. Cummings, and J. Marx, presented at the Physics Education Research Conference*, 2002.

M. S. Sabella, K. Coble, S. P. Bowen, C. Henderson, M. Sabella, and L. Hsu. Using the resources of the student at the urban, comprehensive university to develop an effective instructional environment. In *AIP Conference Proceedings*, volume 1064, page 42, 2008.

K. Samuelowicz and J. D. Bain. Revisiting academics' beliefs about teaching and learning. *Higher Education*, 41(3):299–325, 2001. ISSN 0018-1560, 1573-174X. doi: 10.1023/A:1004130031247. URL http://link.springer.com/article/10.1023/A:1004130031247.

N. E. Sanders, S. Kohler, E. Newton, and Others. Preparing undergraduates for research careers: using Astrobites in the classroom. *Astronomy Education Review*, 11(1), 2012.

W. Sandoval. Conjecture Mapping: An Approach to Systematic Educational Design Research. *Journal of the Learning Sciences*, 23(1):18–36, mar 2014. URL http://www.tandfonline.com/doi/abs/10.1080/10508406.2013.778204.

D. Sawada, M. D. Piburn, E. Judson, J. Turley, K. Falconer, R. Benford, and I. Bloom. Measuring Reform Practices in Science and Mathematics Classrooms: The Reformed Teaching Observation Protocol. *School Science and Mathematics*, 102(6):245–253, oct 2002. ISSN 00366803. doi: 10.1111/j.1949-8594.2002.tb17883.x. URL http://doi.wiley.com/10.1111/j.1949-8594.2002.tb17883.x.

V. Sawtelle, E. Brewe, and L. H. Kramer. Exploring the relationship between self-efficacy and retention in introductory physics. *Journal of research in science teaching*, 49(9):1096–1121, 2012.

R. E. Scherr and R. M. Goertzen. Periscope: Supporting novice university physics instructors in looking into learning in best-practices physics classrooms. In C. Sandifer and E. Brewe, editors, *Recruiting and Education Future Physics Teachers*, pages 215–222. American Physical Society, College Park, MD, 2015.

R. E. Scherr and D. Hammer. Student Behavior and Epistemological Framing: Examples from Collaborative Active-Learning Activities in Physics. *Cognition and Instruction*, 27(2):147–174, apr 2009. ISSN 0737-0008. doi: 10.1080/07370000902797379. URL http://www.tandfonline.com/doi/abs/10.1080/07370000902797379.

W. M. Schlingman, E. E. Prather, C. S. Wallace, A. L. Rudolph, and G. Brissenden. A classical test theory analysis of the Light and Spectroscopy Concept Inventory national study data set. *Astronomy Education Review*, 11:10107, 2012.

A. H. Schoenfeld. What???s All the Fuss About Metacognitlon. *Cognitive science and mathematics education*, 189, 1987.

D. L. Schwartz and J. D. Bransford. A Time For Telling. *Cognition and Instruction*, 16(4):475–5223, dec 1998. ISSN 0737-0008. doi: 10.1207/s1532690xci1604{_}4. URL http://www.tandfonline.com/doi/abs/10.1207/s1532690xci1604{_}4?journalCode=hcgi20.

P. H. Scott, E. F. Mortimer, and O. G. Aguiar. The tension between authoritative and dialogic discourse: A fundamental characteristic of meaning making interactions in high school science lessons. *Science Education*, 90(4):605–631, jul 2006. ISSN 0036-8326. doi: 10.1002/sce.20131. URL http://doi.wiley.com/10.1002/sce.20131.

S. Secules, A. Gupta, A. Elby, and C. Turpen. "Turning Away" from an Undergraduate Programming Student: Revealing Culture in the Construction of Engineering Ability. In *American Educational Research Association 2016 Conference Proceedings*, 2016.

S. B. Seidel and K. D. Tanner. "What if students revolt?"???considering student resistance: origins, options, and opportunities for investigation. *CBE-Life Sciences Education*, 12(4):586–595, 2013.

E. Seymour. *Talking about leaving: Why undergraduates leave the sciences*. Westview Press, 2000.

M. G. Sherin and E. A. van Es. Effects of Video Club Participation on Teachers' Professional Vision. *Journal of Teacher Education*, 60(1):20–37, nov 2008. ISSN 0022-4871. doi: 10.1177/0022487108328155. URL http://jte.sagepub.com/content/60/1/20.short.

E. R. Singer. Espoused teaching paradigms of college faculty. *Research in Higher Education*, 37(6):659–679, dec 1996. ISSN 0361-0365. doi: 10.1007/BF01792951. URL http://link.springer.com/10.1007/BF01792951.

S. R. Singer, N. R. Nielsen, H. A. Schweingruber, and Others. *Discipline-based education research: understanding and improving learning in undergraduate science and engineering*. National Academies Press, 2012.

S. J. Slater. The development and validation of the Test Of Astronomy STandards (TOAST). *Journal of Astronomy & Earth Sciences Education (JAESE)*, 1(1):1–22, 2015.

S. J. Slater, T. F. Slater, and D. J. Lyons. *Engaging in astronomical inquiry*. Macmillan Higher Education, 2010.

T. Slater, J. P. Adams, G. Brissenden, and D. Duncan. What topics are taught in introductory astronomy courses? *The Physics Teacher*, 39(1):52–55, 2001.

M. K. Smith, F. H. M. Jones, S. L. Gilbert, and C. E. Wieman. The Classroom Observation Protocol for Undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices. *CBE life sciences education*, 12(4):618–27, jan 2013. ISSN 1931-7913. doi: 10.1187/cbe.13-08-0154. URL http://www.lifescied.org/content/12/4/618.short.

M. K. Smith, E. L. Vinson, J. A. Smith, J. D. Lewin, and M. R. Stetzer. A Campus-Wide Study of STEM Courses: New Perspectives on Teaching Practices and Perceptions. *CBE-Life Sciences Education*, 13(4):624–635, 2014. doi: 10.1187/cbe.14-06-0108. URL http://www.lifescied.org/content/13/4/624.abstract.

J. P. Smith III, A. A. Disessa, and J. Roschelle. Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, 3(2):115–163, 1994.

D. R. Sokoloff and R. K. Thornton. Using interactive lecture demonstrations to create an active learning environment. In *The changing role of physics departments in modern universities*, volume 399, pages 1061–1074. AIP Publishing, 1997.

J. B. Stang and I. Roll. Interactions between teaching assistants and students boost engagement in physics labs. *Physical Review Special Topics - Physics Education Research*, 10(2):020117, sep 2014. ISSN 1554-9178. doi: 10.1103/PhysRevSTPER.10.020117. URL http://journals.aps.org/prper/abstract/10.1103/PhysRevSTPER.10.020117.

C. M. Steele. A threat in the air: How stereotypes shape intellectual identity and performance. *The American psychologist*, 52(6):613–629, 1997.

C. M. Steele. *Whistling Vivaldi: How Stereotypes Affect Us and What We Can Do*. WW Norton & Co, 2010.

D. Stroupe. *"Students drive where I go next": Ambitious practice, beginning teacher learning, and classroom epistemic communities*. PhD thesis, University of Washington, jul 2013. URL https://digital.lib.washington.edu:443/researchworks/handle/1773/23619.

D. Tannen. *Framing in discourse*. Oxford University Press on Demand, 1993.

The Design-Based Research Collective. Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, pages 5–8, 2003.

A. Traxler and E. Brewe. Equity investigation of attitudinal shifts in introductory physics. *Physical Review Special Topics-Physics Education Research*, 11(2):20132, 2015.

A. L. Traxler, X. C. Cid, J. Blue, and R. Barthelemy. Enriching gender in PER: A binary past and a complex future. *arXiv preprint arXiv:1507.05107*, 2015.

K. Trigwell and M. Prosser. Changing approaches to teaching: A relational perspective. *Studies in Higher Education*, 21(3):275–284, 1996. ISSN 0307-5079. doi: 10.1080/03075079612331381211.

C. Turpen and N. D. Finkelstein. Not all interactive engagement is the same: Variations in physics professors' implementation of Peer Instruction. *Physical Review Special Topics - Physics Education Research*, 5(2):020101, aug 2009. ISSN 1554-9178. doi: 10.1103/PhysRevSTPER.5.020101. URL http://journals.aps.org/prper/abstract/10.1103/PhysRevSTPER.5.020101.

C. Turpen and N. D. Finkelstein. The construction of different classroom norms during Peer Instruction: Students perceive differences. *Physical Review Special Topics - Physics Education Research*, 6(2):020123, nov 2010. ISSN 1554-9178. doi: 10.1103/PhysRevSTPER.6.020123. URL http://journals.aps.org/prper/abstract/10.1103/PhysRevSTPER.6.020123.

C. Turpen, M. Dancy, and C. Henderson. Perceived affordances and constraints regarding instructors' use of Peer Instruction: Implications for promoting instructional change. *Physical Review - Physics Education Research*, 12:010116, 2016.

M. Urry. Speeding up the long slow path to change. *APS News*, 12(2):12, 2003.

L. Valli and W. D. Hawley. Designing and implementing school-based professional development. *The keys to effective schools: Educational reform as continuous improvement*, pages 86–96, 2002.

E. A. van Es and M. G. Sherin. Learning to Notice: Scaffolding New Teachers??? Interpretations of Classroom Interactions. *Journal of Technology and Teacher Education*, 10(4):571–596, 2002. ISSN 1059-7069. URL /p/9171/.

E. A. van Es and M. G. Sherin. Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education*, 24(2):244–276, feb 2008. ISSN 0742051X. doi: 10.1016/j.tate.2006.11.005. URL http://www.sciencedirect.com/science/article/pii/S0742051X06001727.

E. A. van Es and M. G. Sherin. The influence of video clubs on teachers' thinking and practice. *Journal of Mathematics Teacher Education*, 13(2):155–176, 2010.

E. van Zee and J. Minstrell. Using Questioning to Guide Student Thinking. *Journal of the Learning Sciences*, 6(2):227–269, apr 1997. ISSN 1050-8406. doi: 10.1207/s15327809jls0602{_}3. URL http://www.tandfonline.com/doi/abs/10.1207/s15327809jls0602{_}3.

L. S. Vygotsky. *Mind in society: The development of higher psychological processes*. Harvard University Press, Cambridge, MA, 1980.

C. Wainwright, P. D. Morrell, L. Flick, and A. Schepige. Observation of Reform Teaching in Undergraduate Level Mathematics and Science Courses. *School Science and Mathematics*, 104(7):322–335, nov 2004. ISSN 00366803. doi: 10.1111/j.1949-8594.2004.tb18251.x. URL http://doi.wiley.com/10.1111/j.1949-8594.2004.tb18251.x.

C. L. Wainwright, L. B. Flick, and P. D. Morrell. Development of instruments for assessment of instructional practices in standards-based teaching. *Journal of Mathematics and Science: Collaborative Explorations*, 6(1):21–46, 2003.

C. Walkington and M. Marder. Classroom Observation and Value-Added Models Give Complementary Information about Quality of Mathematics Teaching, 2013.

C. S. Wallace, E. E. Prather, and B. M. Mendelsohn. Astro 101 Students' Perceptions of Science: Results from the Thinking About Science Survey Instrument. *Astronomy Education Review*, 12(1):10101, 2013.

G. C. Weaver, W. D. Burgess, A. L. Childress, and L. Slakey, editors. *Transforming Institutions: Undergraduate STEM Education for the 21st Century*. Purdue University Press, West Lafeyette, Indiana, 2016.

E. A. West, C. A. Paul, D. Webb, and W. H. Potter. Variation of instructor-student interactions in an introductory interactive physics course. *Physical Review Special Topics - Physics Education Research*, 9(1):010109, mar 2013. ISSN 1554-9178. doi: 10.1103/PhysRevSTPER.9.010109. URL http://journals.aps.org/prper/abstract/10.1103/PhysRevSTPER.9.010109.

C. Wieman, K. Perkins, and S. Gilbert. Transforming Science Education at Large Research Universities: A Case Study in Progressxs. *Change: The Magazine of Higher Learning*, 42(2):6–14, 2010.

B. R. Wilcox, B. M. Zwickl, R. D. Hobbs, J. M. Aiken, N. M. Welch, and H. J. Lewandowski. Alternative model for assessment administration and analysis: An example from the E-CLASS. *arXiv preprint arXiv:1601.07896*, 2016.

K. E. Williamson and S. Willoughby. Student understanding of gravity in introductory college astronomy. *Astronomy Education Review*, 11(1):10105, 2012.

S. M. Wilson. Professional development for science teachers. *Science*, 340(6130): 310–313, apr 2013. ISSN 1095-9203. doi: 10.1126/science.1230725. URL http://science.sciencemag.org/content/340/6130/310.abstract.

B. Zwickl, N. D. Finkelstein, and H. J. Lewandowski. Development and Validation of the Colorado Learning Attitudes about Science Survey for Experimental Physics. In *Physics Education Research Conference 2012*, pages 442–445, Philadelphia, PA, 2012. PER Conference.