Instructor interactions in traditional and nontraditional labs

David G. Wu
Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA

Ashley B. Heim
Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853, USA

Meagan Sundstrom, Cole Walsh, and N. G. Holmes
Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA

(Received 1 December 2021; accepted 18 January 2022; published 14 March 2022)

As physics laboratory courses (labs) transition from traditional, model-verifying activities to discovery-based investigations, it becomes crucial to understand the role of the instructor in the implementation of various lab types. Prior work has started to address this need by examining either coarse-grained frequencies or fine-grained content of instructor interactions in labs. However, neither of these methods offer both a detailed and time-efficient procedure for measuring such interactions, which is required for comparisons across multiple sessions of several types of labs. Here we describe and present the results of a new approach to quantifying student-instructor interactions in labs by analyzing video recordings and drawing on techniques of social network analysis. Across five sections of three lab course offerings, we find that there is a higher total level of interaction between students and instructors in reformed discovery-based labs than in traditional labs. We find no clear pattern in the durations of student-initiated and instructor-initiated interactions across various instructors and lab types. The results suggest that the amount of interaction between instructors and students during lab is more a product of the instructional design than an individual instructor's implementation of that design. This work is a preliminary step toward understanding the extent to which student-instructor interactions support the improved outcomes observed in discovery-based labs compared with traditional labs.

DOI: 10.1103/PhysRevPhysEducRes.18.010121

I. INTRODUCTION

In recent years, many introductory physics laboratories (labs) have been transitioning from traditional, model-verification formats to more discovery-based or open-ended formats. Many of the studies evaluating lab instruction have focused on student outcomes and student behaviors. For example, compared to traditional, model-verification methods, discovery-based or open-ended labs (hereby referred to as *reformed* labs) have been found to improve student attitudes and beliefs towards experimentation [1–3] and to increase student engagement in experimentation skills and abilities [3–8]. Few studies, however, have identified the instructional mechanisms responsible for these improved outcomes. In this study, we evaluate the role of instructor interactions in labs by comparing between three different instructional lab offerings.

As lab curricula evolve, the role of the instructor is key to the success of the implementation. Students in labs typically work in small groups, engaging in more student-to-student and student-to-instructor interactions compared to lectures [9]. At larger universities, furthermore, laboratory courses are often taught by graduate teaching assistants (TAs) rather than faculty. The TAs are rarely involved in the instructional design of the lab and so are serving as agents of the lead instructor's instructional mission. The ways the TAs understand and enact the instructional goals are, therefore, pivotal for the potential success of the course. For example, the frequency with which an instructor interacts with students in labs directly impacts their engagement [10,11]. Different instructors, however, may interact with students in different ways, ultimately implementing the same instructional materials quite differently [11–16]. This is particularly true for reformed labs, where the instruction supports responsive teaching and teacher agency [13].

In evaluating the role of the instructor, both the content and frequency of the instructor's interactions may impact student engagement and learning. Physics education researchers have used multiple means of measuring instructors' behaviors in lab, which range from coarse-grained systematic observations to more fine-grained qualitative analyses.

Observational tools such as the Laboratory Observation Protocol for Undergraduate STEM (LOPUS) [17] and the Real-time Instructor Observing Tool (RIOT) [18] systematically document the instructor's and students' activities throughout the lab. The LOPUS has been used to characterize the frequency and content of instructor's interactions with student groups in labs [9,11,17,19]. Observers using the LOPUS identify at regular time intervals whether the instructor is engaging in typical instructional behaviors (e.g., lecturing to the class, writing on the board, presenting a demonstration, or monitoring the class), more interactive behaviors (e.g., talking or posing questions to individual groups or providing positive reinforcement), or engaging in noninstructive behaviors (e.g., performing administrative tasks or not interacting with the students). The content of these interactions is also characterized under the LOPUS as either discussing underlying scientific principles, data analysis and calculations, experimental procedures, safety, or previous material. Similarly, the RIOT has been used to characterize the kinds of instructional activities the instructor performs throughout the lab [13,20,21]. Observers using the RIOT continuously identify whether the instructor is interacting with the whole class, an individual student, or an individual group (and which group), or simply observing the class. The content of their interactions is characterized as either clarifying instructions, explaining content, listening to student questions, or engaging in open or closed dialogue. These observational protocols typically require an observer to be present in the lab room and to be able to hear the instructor's interactions with all students.

On a finer-grain scale, researchers have also used video and audio recordings of instructors in labs to characterize their interactions with students in more detail. For example, researchers may perform content analysis on all instructor talk throughout the lab [22]. The TA Practices In and Views Of Teaching framework (TA-PIVOT [14]) uses interviews and recorded observations to characterize instructor's beliefs and practices during instruction.

Each of these approaches provides detailed information about the ways in which instructors interact with students in labs. However, with increased detail comes decreased efficiency. For example, content analysis and the TA-PIVOT are suitable for case studies, given the researcher time investment required to analyze instructor and student speech with these two methods. Observation protocols such as the LOPUS and the RIOT are more efficient and coarse-grained, but still require a researcher to listen to the instructor's interactions during the full lab sessions (typically two hours each). Characterizing interactions for a large number of lab sessions in multiple courses requires a coarser grained, holistic understanding of instructor-student interactions.

Fortunately, research has shown that simple measures of the frequency with which the instructor interacts with students are correlated with students' behaviors [10,11]. One might expect that students' conversations with the instructor are more likely to be on-topic than conversations between peers and so it is easier to infer that interactions with the instructor help keep students on task. Interestingly, one research study found that instructor-initiated interactions, but not student-initiated interactions, were found to promote student engagement, regardless of the length of those interactions [10]. They argued that the instructor-initiated interactions improve engagement through a policing effect, though other research has argued for a sense of immediacy, whereby the closeness of the instructor (physically and, because the instructor opens the conversation, psychologically) encourages students to engage with them [23]. This sense of immediacy was shown to relate to the instructor's behaviors in the class and correlated positively with students' affective and cognitive outcomes [23]. While students may or may not initiate interactions based on a number of factors, instructor-initiated interactions are often more uniform and constant in the classroom.

In this study, we sought to evaluate the interactions that graduate TAs and instructors (generally referred to as *instructors* from here) have with students in traditional and reformed labs. To efficiently compare across multiple lab sessions in multiple lab courses, we used video data to identify when the instructor was interacting with individual groups and with the whole class. We used a combination of the instructor's position in the room and audio to confirm the start and end of interactions. Using techniques from social network analysis, we compare the interactions between instructors and students in the two types of labs, including two iterations of a reformed lab, as well as among different instructors. The analysis takes into account the duration of interactions, who initiated these interactions, and who is involved in the interactions.

We find that there is a higher level of interaction between students and instructors in reformed labs than in traditional labs. We also observe the same amount of student-instructor interaction in labs of the same type for a given lab session, even with different instructors, students, and amount of student-initiated and instructor-initiated interactions. Surprisingly, we find no clear pattern between the levels of instructor-initiated or student-initiated interactions between the courses or between instructors. The results suggest that the variability between the instructional lab curricula likely impact instructor behavior (and, in turn, student outcomes) more than the instructors' individual enactment of the lab curricula. This work is a preliminary step towards understanding the degree to which the students' interactions with the instructor support the improved outcomes seen in discovery-based or open-ended lab curricula compared with traditional labs.

TABLE I. Summary of information about the lab type, instructors, and number of codable sessions, as well as the self-reported major and year of students in each lab section. Numbers in parentheses are the N values corresponding to the percentages.

Lab information		Section 1	Section 2	Section 3	Section 4	Section 5
	Lab instruction type	Traditional	Old reformed	Old reformed	New reformed	New reformed
	Instructor	Teresa	Owen	Oliver	Noah	Noah
	Codable sessions	7	8	8	8	7
Self-reported student information						
Race or Ethnicity	Asian or Asian American	54% (13)	22% (5)	28% (7)	32% (6)	29% (4)
	Black or African American	4% (1)	0	0	5% (1)	0
	Hispanic or Latino	0	4% (1)	12% (3)	16% (3)	7% (1)
	White or Caucasian	33% (8)	70% (16)	52% (13)	26% (5)	57% (8)
	Multiple races or ethnicities	8% (2)	0	8% (2)	16% (3)	7% (1)
	Other or unknown	0	4% (1)	0	5% (1)	0
Gender	Male	71% (17)	74% (17)	84% (21)	58% (11)	36% (5)
	Female	29% (7)	26% (6)	16% (4)	42% (8)	64% (9)
Major	Physics or Engineering Physics	58% (14)	39% (9)	36% (9)	0	0
	Other Engineering or Computer Science	13% (3)	30% (7)	24% (6)	69% (13)	57% (8)
	Other STEM	4% (1)	18% (4)	20% (5)	26% (5)	36% (5)
	Unknown	25% (6)	13% (3)	20% (5)	5% (1)	7% (1)
Year	First year	100% (24)	96% (22)	96% (24)	53% (10)	71% (10)
	Second year	0	0	4% (1)	42% (8)	29% (4)
	Third year	0	4% (1)	0	5% (1)	0

II. METHODS

The data for this study come from five laboratory sections of introductory mechanics courses at Cornell University. Information about the lab type, instructor, number of codable lab sessions, and student composition of each lab section are summarized in Table I. The instructors of these labs consisted of four different individuals: three men and one woman. Three were experienced graduate teaching assistants (one with a background in physics education research) and one was a faculty member in the physics department. To protect the anonymity of these instructors, specific information regarding which instructor taught which lab section will not be provided. We will use pseudonyms for instructor names, with the first letter of each name corresponding to the lab type each instructor taught. The topics explored in each lab session had been introduced to students in class prior to each lab session, with the topics of each session for each section summarized in Table II. More detail about these labs can be found in our other work [3,24,25].

A. Lab instruction types

Section 1 was a *traditional* lab from an introductory mechanics course designed for first-year physics students. In these traditional labs, students worked in groups of two to four students and followed a lab manual to confirm a

TABLE II. Topics of lab activities for each session for each lab type. The project labs were sessions where students had the freedom to explore a question building off one of the previously studied topics. The old reformed labs had an extra lab session studying "collisions" between sessions six and seven that the new reformed labs did not. This session was not codable and was omitted from all analysis, so the session numbers for the old reformed labs are modified to match those in the new reformed labs.

Session	Traditional (Section 1)	Old reformed (Section 2 and 3)	New reformed (Sections 4 and 5)
1	Dynamics of 1D Motion	Pendulum I	Pendulum I
2	Newton's Laws and Dynamics	Pendulum II	Pendulum II
3	Force Laws	Terminal Velocity I	Terminal Velocity I
4	Energy Exchanges	Terminal Velocity II	Terminal Velocity II
5	Collisions I	Hooke's Law I	Hooke's Law I
6	Collisions II	Hooke's Law II	Hooke's Law II
7	Angular Momentum	Project Lab I	Project Lab I
8	Oscillations	Project Lab II	Project Lab II

model or to explain a particular phenomenon. Over the semester, there were eight lab sessions, each scheduled for two hours. However, students in the traditional labs were able to leave once they had finished their lab assignment, so each session lasted around an hour and fifteen minutes, on average. Teresa was the instructor of this traditional lab section.

Sections 2 and 3 were *old reformed* labs and came from the same introductory mechanics course designed for first-year physics students and in the same semester as section 1. These reformed labs strongly emphasized the development of critical thinking and experimental physics skills through open-ended investigations. In each session, students worked together in groups of two to three students to design and conduct experiments revolving around a particular topic or concept. The instruction was designed to increasingly foster student agency, leading to a student-driven project lab at the end of the semester [3,25]. Unlike the traditional labs, students stayed for the whole lab period, so each lab session lasted around two hours. Sections 2 and 3 were taught by Owen and Oliver, respectively.

Sections 4 and 5 were *new reformed* labs from an introductory mechanics course designed for first-year engineering students. Although these sections were also reformed labs with activities similar to those in sections 2 and 3, sections 4 and 5 were implemented two years later and are thus referred to as new reformed labs. There were slight differences to the instruction and curriculum due to changes over time, with minor improvements made to some sessions. The new reformed labs were also implemented with a new student population, such that students were primarily engineering, rather than physics, majors (Table I). The new reformed sections had eight lab sessions, each roughly two hours long where students stayed for the full duration. Unlike the old reformed sections, both new reformed sections were taught by the same instructor, Noah.

B. Coding

Video and audio data of each lab session were obtained through wall-mounted cameras in the lab classrooms and a portable microphone worn by the lab instructor. These data allowed us to observe interactions between the lab instructor and each group, as well as to hear the discourse between groups during whole-class discussion. Most lab sections had one session that was not codable, either due to an obstructed camera or technical issues. This left seven or eight codable sessions for each section.

Each codable lab session was coded in BORIS, a behavioral analysis software that allows for the easy logging of social interactions [26]. Researchers coded the duration and directionality (who initiated each interaction) of each interaction between the instructor and students. There were two types of interactions: *group* interactions and *whole-class* interactions.

The beginning of a group interaction was defined by audible speech between the instructor and an individual lab group. The end of these interactions was signaled by the end of the conversation, often accompanied by the instructor physically leaving the area. Brief exchanges between a group and the instructor (that lasted fewer than three seconds) and that did not hold substantial content (such as a quick yes or no question) were not recorded as interactions.

Whole-class interactions included interactions where the instructor addressed the whole class, such as for lecturing or to give instructions or advice. The beginning and end of these interactions were clearly defined by the moment when the instructor began and stopped talking to the whole class. Interactions between an individual group and the whole class were also recorded, such as when students gave short presentations to the whole class about their experiments and findings or when students gave a substantial answer (not just a yes or no answer) to a question posed by the instructor during whole-class discussion. The beginning and end of these interactions were similarly defined by the moments when the students began and stopped talking.

The directionality of the interactions (who initiated the interaction) fell into two categories: *student-initiated* interactions and *instructor-initiated* interactions. The initiator of an interaction was often characterized by who spoke first. However, instances where a student raised their hand (regardless of who spoke first) were considered student-initiated interactions. Similarly, instances where the instructor intentionally approached a lab group but a student within the group spoke first were considered instructor-initiated interactions.

The resulting data provide the duration and directionality of all interactions between groups and the instructor as well as interactions during whole-class discussion. Using this coding scheme, the coders were able to watch video from each lab session at a fast-forwarded rate, recording the beginning, ending, and directionality of all interactions for each roughly two-hour lab session in about an hour. This method is much more efficient than both the observation protocols and qualitative analyses described in the introduction, facilitating large-scale data analysis across multiple lab sessions and lab sections. This method fails, however, to capture the content of the interactions, providing breadth, but not depth, of instructor interactions.

The five sections were independently coded by four of the authors. To gauge the reliability of this method, all four coders coded the same lab session. We calculated edge and weight agreement between pairs of coders [27]. Edge agreement is the number of edges coded by a pair of coders as a fraction of the number of edges coded by at least one of the two coders. Weight agreement is the sum of edge weights (time duration of interactions) coded by a pair of

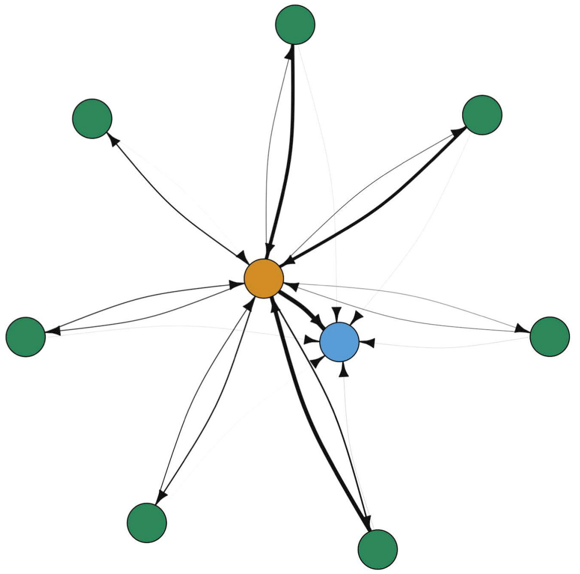


FIG. 1. Sample network diagram. The central, orange node corresponds to the instructor, the green nodes correspond to a student group, and the blue node corresponds to the whole class (so that edges between an orange or green node and the blue node represent presentations to the whole class). The edges connecting the nodes are weighted to represent the duration of the interactions and directional to indicate who initiated the interaction.

coders as a fraction of the sum of edge weights coded by at least one of the two coders. As shown in Ref. [28], these measures are quite volatile and serve as conservative estimates of interrater reliability [27]. Unfortunately, there is no standard threshold for reliability with network data. In this study, edge agreement ranged from 68%–95% (with an average of 82%) and weight agreement ranged from 51%–86% (with an average agreement of 69%). As in previous work [28], our networks with higher density are more reliable than sparse networks.

C. Analysis

The coded data were converted to networks of interactions between the instructor and the student groups (as nodes) in each session. The whole class interactions were added to the networks as a unique node (rather than as edges from the speaker to each other node). The nodes were connected by edges that represented whether the instructor interacted with that group (or the whole class). The edges were directional based on who initiated the interaction and weighted based on the total duration of the interaction. See Fig. 1 for an example of a network for one lab session.

The coded data were then summarized using the network measure of *strength* [29] (Table III). In social network analysis, strength is defined as the weighted total number of edges attached to an individual node. In our work, we measure the strength for the node representing the instructor, corresponding to the total time the instructor spent interacting with the groups in a particular lab session. Because the duration of each lab sessions differed slightly (particularly in the traditional lab section), we normalized the strength by the total duration of each lab session, marked as the time between when the instructor started class to when the last student left the room. We separately calculated the strength of the instructor for interactions with individual groups, with the whole class, and with groups plus the whole class, and differentiated strength for student-initiated and instructor-initiated interactions (Table III).

First, $total\ strength_{\text{group}}$ measures the proportion of time that the instructor interacted with individual groups (excluding whole-class interactions). $Instrength_{\text{group}}$ measures the percentage of each lab session the instructor spent on student-initiated group interactions, while $outstrength_{\text{group}}$ measures the percentage of each lab session the instructor spent on instructor-initiated group interactions. For a particular session, the sum of the

TABLE III. Summary of all measures used to evaluate interactions.

Measure	Description
Instrength$_{\text{group}}$	The percentage of time in a particular lab session spent on student-initiated interactions between the instructor and individual groups.
Outstrength$_{\text{group}}$	The percentage of time in a particular lab session spent on instructor-initiated interactions between the instructor and individual groups.
Total Strength$_{\text{group}}$	The percentage of time in a particular lab session spent on all interactions between the instructor and individual groups. Total strength$_{\text{group}}$ is equal to the sum of instrength$_{\text{group}}$ and outstrength$_{\text{group}}$.
Total Strength$_{\text{group+whole class}}$	The percentage of time spent in a particular lab session on all interactions between the instructor and individual groups plus interactions between the instructor and the whole class.
Total interaction percentage	The percentage of time spent in a particular lab session on all interactions between the instructor and individual groups, interactions between the instructor and the whole class, and interactions between an individual group and the whole class.

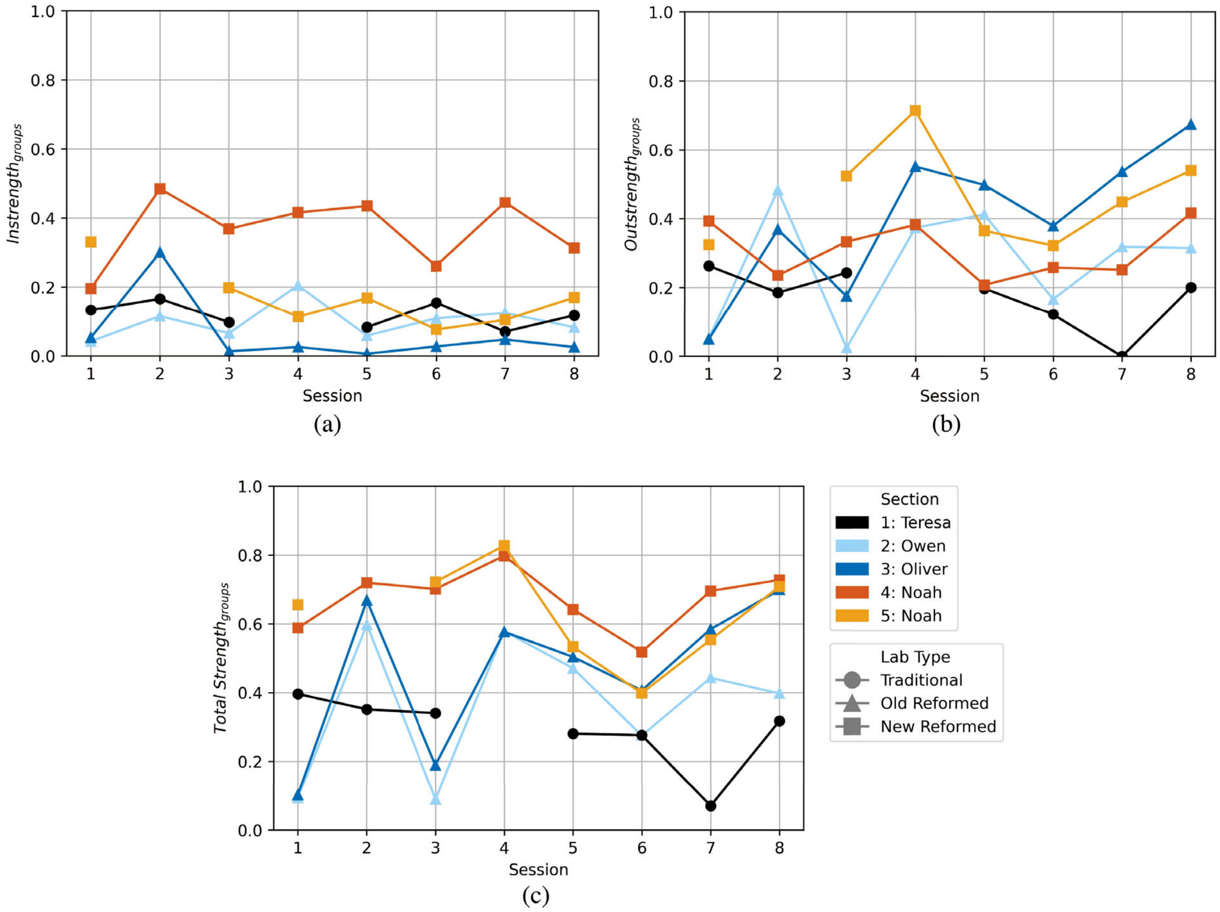


FIG. 2. Strength measures for interactions between the instructor and each individual group normalized by the duration of the lab session. Missing data points correspond to uncodable lab sessions. Instrength$_{\mathrm{group}}$ includes only student-initiated interactions, outstrength$_{\mathrm{group}}$ includes only instructor-initiated interactions, and total strength$_{\mathrm{group}}$ includes all interactions. (a) Instrength$_{\mathrm{group}}$ for all codable sessions. (b) Outstrength$_{\mathrm{group}}$ for all codable sessions. (c) Total strength$_{\mathrm{group}}$ for all codable sessions.

instrength$_{\mathrm{group}}$ and outstrength$_{\mathrm{group}}$ is equivalent to the total strength$_{\mathrm{group}}$.

We also calculated *total strength*$_{\mathrm{group+whole\,class}}$, which measures the proportion of time the instructor spent interacting with individual groups plus whole-class interactions. Thus, this measure is the same as total strength$_{\mathrm{group}}$ plus the proportion of time the instructor addressed the whole class. Note that this measure does *not* include interactions between an individual group and the whole class since these interactions do not explicitly involve the instructor. Instead, we separately calculated the *total interaction percentage*, which is the proportion of the lab period that is spent on interactions between groups and the instructor, between the instructor and the whole class, and between a group and the whole class. Interactions between a group and the whole class were included in this measure because the instructor is also engaged in that interaction. Thus, the total interaction percentage is the total time the instructor is engaged with the students.

These network metrics do not have defined methods for determining uncertainties, meaning we cannot use statistical methods of comparing the values between sessions or sections. In the results, therefore, we look for general patterns and trends in the data, rather than comparing individual points quantitatively.

III. RESULTS

A. Group interaction measures

Figure 2 summarizes the interactions between the instructor and the student groups. Figure 2(a) shows the instrength$_{\mathrm{group}}$ of the instructor for each lab session for all five lab sections across the duration of the semester, which corresponds to the proportion of time the instructor spent in student-initiated group interactions. The instructors in all sections have a relatively stable instrength$_{\mathrm{group}}$, with Oliver's section and Noah's section 4 having the largest range of about 30% over the semester. Additionally, Teresa, Owen, Oliver, and Noah in section 5 have comparable instrength$_{\mathrm{group}}$, with all but three of thirty lab sessions residing within the 0%–20% range. This indicates that the portion of the lab taken up by student-initiated interactions is similar across these sections. Noah in section 4, however, consistently has a higher instrength$_{group}$ than the other instructors (including his own other section). Noah's section 4 instrength$_{group}$ resides primarily within the 30%–50% range, showing that a greater portion of time was spent on student-initiated interactions in this section than in the other sections. There are also no apparent trends over time or among labs of the same type: each instructor's instrength$_{group}$ takes its own unique trajectory through the semester.

Figure 2(b) plots the outstrength$_{group}$ of the instructor for all lab sessions, which corresponds to the proportion of time the instructor spent in instructor-initiated group interactions. This measure is much noisier than instrength$_{group}$, exemplified particularly by Oliver's large variability (range of 62%). No single section stands out from the others as particularly high or particularly low (as compared with the systematically higher instrength$_{group}$ for Noah's section 4). Similar to instrength$_{group}$, outstrength$_{group}$ does not show any consistent trends over time or between sections of the same lab type.

Unlike instrength$_{group}$ and outstrength$_{group}$, total strength$_{group}$ shows strong trends between sections of the same lab type, shown in Fig. 2(c). Instructors in sections of the same lab type have nearly identical measures of total strength$_{group}$ for any particular session, despite having different instrength$_{group}$ or outstrength$_{group}$. Sections 4 and 5, from the new reformed labs, map almost directly onto one another throughout the course of the semester and consistently hold the highest total strength$_{group}$ compared with the other lab types. Sections 2 and 3, from the old reformed labs, map each other closely and typically have a lower total strength$_{group}$ than the new reformed labs for any given week. Furthermore, the trend of the interactions from session four onwards (decreasing from four to six and then increasing) is similar across all four reformed lab sections (except for section 2's eighth session, where there was a disruptive incident outside the building which could be seen from the window of the lab classroom, changing the nature of many social interactions [30]). Interestingly, the lab activities in these final sessions were the most similar between the new and old reformed labs.

Section 1, from the traditional labs, has, on average, some of the lowest values of total strength$_{group}$ compared with the other sections. Interestingly, the similarities in total strength$_{group}$ between sections of the same lab type appear to hold even with different instructors. This is evidenced by the similarities in total strength$_{group}$ for sections 2 and 3, which were taught by different instructors (Owen and Oliver, respectively).

Despite seeing no apparent patterns between sections of the same lab type in terms of instrength$_{group}$ and outstrength$_{group}$, the addition of these two measures and the resulting total strength$_{group}$ shows that the total time the instructor spends talking to individual groups during a particular lab session is similar between sections of the same lab type. The lower pattern of total strength$_{group}$ for Teresa's section matches our expectations that the traditional labs were not as conducive to student-instructor interactions as the reformed labs. The tracking fluctuations in the total strength$_{group}$ within lab types will be discussed further below.

B. Whole-class interaction measures

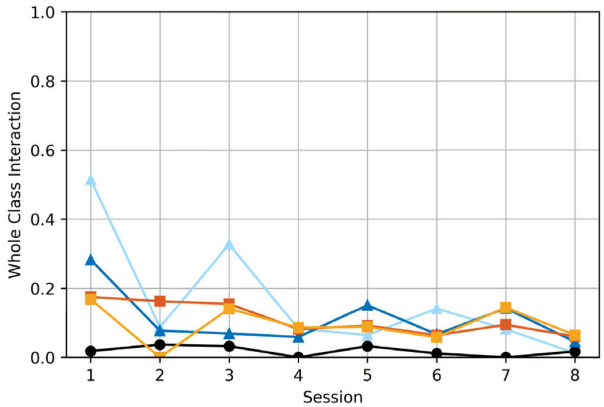
Figure 3 summarizes the instructor interactions including the whole-class interactions. Figure 3(a) shows the percentage of time the instructor spent leading a whole-class discussion. Other than sessions one and three, instructors in the reformed labs spent less than 20% of class time in whole-class interactions. The traditional lab section (section 1) spent little to no time in whole-class interactions in any of the sessions. We see that the old reformed labs (sections 2 and 3) had spikes in whole-class interactions in sessions one and three, which correspond to dips in the total strength$_{group}$ (time spent interacting with student groups).

Figure 3(b) shows the total strength$_{group+whole\,class}$, which represents the total proportion of each lab session the instructor spent interacting with groups through both individual group and whole class interactions. Just like with total strength$_{group}$, total strength$_{group+whole\,class}$ illustrates clear similarities between sections of the same lab type. Noah consistently has the highest total strength$_{group+whole\,class}$ in both sections 4 and 5. Owen and Oliver's total strength$_{group+whole\,class}$ sit below the new reformed labs, with their total strength$_{group+whole\,class}$ measures mapping almost directly onto one another (with the exception of sessions seven and eight). With the addition of discussion between the instructor and the whole class, the low values in sessions one and three of the old reformed labs are less pronounced in Fig. 3(b) and more consistent with the rest of the sessions.

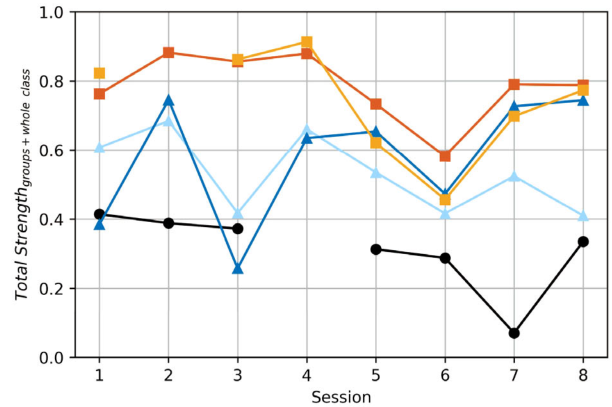
As with total strength$_{group}$, Teresa consistently has one of the lowest total strength$_{group+whole\,class}$. Notably, Teresa's total strength$_{group+whole\,class}$ is not much different than her total strength$_{group}$, reflective of the lack of discussion between the instructor and the whole class in the traditional labs shown in Fig. 3(a). This was expected given that groups in this section had to complete procedural lab activities that did not necessitate whole class discussions. In contrast, whole-class discussions were designed into the reformed lab curriculum.

C. Total interaction percentage

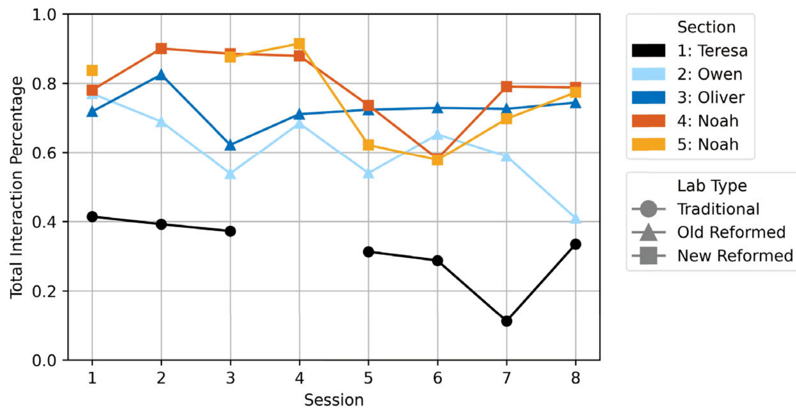
Finally, Fig. 3(c) shows the proportion of each lab session the instructor spent engaged in any type of interaction (the total interaction percentage). This is the proportion of time the instructor spent interacting with groups individually and as a whole class, as well as



(a) Percentage of interactions between the instructor and the whole class for all codable sessions.



(b) *Total strength*$_{group+whole\ class}$ for all codable sessions.



(c) *Total interaction percentage* for all codable sessions.

FIG. 3. Strength measures for (a) duration of instructor and whole class interactions only, (b) duration of instructor interactions with the whole class or individual groups, and (c) total duration of all instructor interactions and interactions involving a group and the whole class, all normalized by the duration of the lab session. Missing data points correspond to uncodable lab sessions.

interactions between individual groups and the whole class. Once again, the total interaction percentage of lab sections belonging to the same lab type map closely to one another. Sections 4 and 5 are almost identical, while sections 2 and 3 track similarly with the exception of session eight when the distracting incident occurred in section 2.

Notably, with the addition of interactions between individual groups and the whole class, all reformed labs, both new and old, begin to overlap. This suggests that both types of reformed labs allow for roughly the same proportion of interactions relative to the duration of the lab period, but the way in which this time is used differs. The new reformed labs spend more time on instructor interactions with the individual groups while the old reformed labs spend more time on whole-class discussion, whether it be between an individual group and the whole class or between the instructor and the whole class. The old reformed labs particularly involved student participation in whole-class discussions during sessions one and three. From the videos, the whole-class discussions during these sessions were associated with lessons on new data analysis tools. While similar lessons were used in the new reformed labs, the students in the old reformed labs more frequently participated in these lessons, such as by sharing their ideas from the activity and answering instructor-posed questions. For the new reformed labs, instructors were explicitly trained to keep these same activities much shorter. The total interaction percentage also solidifies that the traditional labs do not facilitate instructor interaction, as Teresa's total interaction percentage is by far the lowest in every session.

IV. DISCUSSION

In this paper, we evaluated the degree to which instructors interact with student groups during instructional physics labs. By comparing across instructors and different types of labs, we find that reformed labs consistently facilitate more interaction between students and the instructor than a traditional lab. Furthermore, we observe that instructors in the same lab type interact with students the same amount during each session, despite differences in student groups and the instructors themselves. Interestingly, we find significant variability in our various measures of interactions between lab sessions.

Together, these results suggest that the instructional design of the labs more strongly impacts the amount of student-instructor interaction than any individual implementation. For example, if the amount of interaction was primarily due to an instructor's initiative, we might expect little to no variability between sessions in the new reformed labs (sections 4 and 5), which were taught by the same instructor. Instead, we find a fair amount of variability across sessions. We would also expect substantial differences in the interactions between sections of the old reformed labs (sections 2 and 3), which were taught by different instructors. Instead, we see that the interactions in these two sections track quite closely until session six. After session six, Oliver's interactions track more closely to those of Noah in the new reformed labs. Interestingly, the lab activities in sessions five through eight were the most similar between the new and reformed labs (fewer changes made to the instructions over time). The difference between Owen's interactions and those of the other instructors during those sessions may or may not be quantitatively meaningful. Because uncertainties are undefined in these data, there is no accepted way to quantitatively compare these data to one another. Thus, we can speak only about patterns and qualitative (rather than statistical) differences between the interactions. The trend of Owen's interactions over those final four sessions follow a similar pattern to those of the other reformed lab instructors, barring the final session, when there was a distracting incident just outside the lab room [30].

Multiple instructional variables may be responsible for the patterns of interactions among lab types. The most salient explanation is that variability between lab types is due to the nature of the lab instruction in these courses, particularly the available inquiry levels [31] or student autonomy [19]. With more open-ended labs, we would expect students have more to discuss with the instructor (to get feedback on the design, interpreting results, overcoming surprising findings, on top of clarifying technical details), thus facilitating more interactions. Our data partly support this interpretation, particularly by comparing interactions *within* the reformed labs. For example, both sections 2 and 3 experience a dip in interactions during session three, when students worked on a terminal velocity lab. While the investigation was set up as one with an ambiguous outcome (students had to collect data to evaluate which of two models for drag best described the falling motion of coffee filters [32]), students experienced significant technical challenges with the detector and, anecdotally, reverted to framing the lab as a more traditional, verification lab. They had little choice in their experimental design and the technical challenges further limited that choice. As a result, this lab, in particular, was significantly redesigned in the new reformed labs (see Ref. [33] for a description and analysis of the redesigned version of the lab), which may explain the much higher levels of interaction in sections 4 and 5 for the corresponding lab.

Alternatively, the amount of interaction may simply be a result of the time available to the instructors to interact with the groups outside of whole-class activities. For example, instructors in course-based undergraduate research experiences (i.e., CUREs) were found to engage in more discussions with students than in traditional labs in part due to reduced time spent lecturing in these courses [19]. This interpretation is supported by comparing the instructor-student interactions as a function of the amount of time spent in whole-class discussions (time when students have no autonomy in their lab work). For example, the percentage of whole-class interactions in section 2 are highest in session one, drop substantially in session two, pick up again in session three, and drop in session four [Fig. 3(a)]. The strength of interactions between the instructor and the individual groups shows the exact opposite pattern: interactions are lowest in session one, high in session two, drop again in session three, and peak again in session four [Fig. 2(c)].

Additional instructional mechanisms beyond whole-class discussion time, however, are also likely at play. For example, sections 4 and 5 (the new reformed labs) show relatively flat percentages of whole-class discussions, despite variability in the instructor-group interactions, particularly towards the end of the semester. The same is true in section 1 (the single traditional section in the dataset), which had both the least time in whole-class discussions and in group interactions. These comparisons again point towards the lab design being primarily responsible for our results, given that the reformed labs intentionally designed whole-class discussions and student presentations into the instruction, while the traditional labs did not.

The differences in total interactions between lab types may also or instead be due to the training the instructors received, which was iteratively improved over time (and therefore between lab types). Research has shown that instructor training significantly impacts instructors' behaviors in instructional labs [9,34,35]. In the new reformed labs, instructors attended weekly workshops focused on pedagogical techniques to support student learning. Instructors in the old reformed labs also met weekly, but the focus of the session was more on what the students would be doing, rather than actions for the instructor. The instructors in the traditional sections did not meet for training at all. One example of where this training may have been impactful is in the students' participation in whole-class discussions in the reformed labs. In both the old and new reformed labs, sessions one and three involved instructor-led data analysis activities. In the old reformed labs, the instructors took time to present on the analysis tools being taught, with students presenting their ideas and responding to instructor-posed questions. In the new reformed labs, instructor training focused explicitly on making the discussion brief so that students had sufficient time to apply the new data analysis tools with their experiments.

Somewhat surprisingly, the proportion of interactions does not seem to correlate with technical challenges in the labs. In the old reformed labs, for example, session three involved the technical challenges described above but included some of the fewest instructor-group interactions. In contrast, session two involved students measuring periods of a pendulum with a stopwatch (technically very straightforward), but included some of the most instructor-group interactions.

Overall, the data do not support attributing the differences between lab types to the individual instructors themselves, given the similarities in total strength of interactions between the two different instructors in sections 2 and 3. The ways in which these instructors interact, however, are different, as demonstrated by the variability in instrength and outstrength between sections. We infer that each lab session (in each type of lab) has a set amount of time that the instructor can take up and how that time is taken up varies between instructors, students, and aspects of the instructional materials.

One choice instructors have in taking up the available time is whether to initiate interactions or respond to student-initiated interactions. We found substantial variability in the proportions of instructor-initiated or student-initiated interactions between sessions, sections, and lab types. For example, consider the two sections with the most student-instructor interactions (sections 4 and 5), both taught by the same instructor with the same curriculum. Section 4 had the most student-initiated interactions (instrength) of all the sections, while section 5 had some of the most instructor-initiated interactions (outstrength). Thus, whether the interactions are student- or instructor-initiated is likely more a function of the students than the instructor. For example, other work analyzing students in these videos found that one group in section 4 exhibited significant help-seeking behaviors, regularly seeking out guidance from the instructor [33]. Other work also found that students in section 4 engaged in more group-group interactions than students in other sections [28], so these students may simply have been more social.

Other data support the role of student independence (i.e., versus student-initiated help-seeking) in the amount of interaction. For example, for sessions five and six in all four reformed lab sections, students explored their own research questions related to testing Hooke's law on everyday objects. Very few changes were made to this lab between the old and new reformed labs. Interestingly, all four reformed sections experience relatively similar levels of interactions with the instructors during these sessions. Furthermore, three of the four sections show particularly low levels of interaction during these sessions, which may reflect students being self-sustaining in their independent investigations. This hypothesis appears to be contradicted by the "project lab" (sessions seven and eight, where students investigated their own research questions about any physics topic). Three of the four sections see a spike in instructor interactions in these sessions, while section 2 sees a dip in session eight (though this may be due to the disruptive incident that occurred during this particular lab session). Although the project lab is designed to have students pose their own research questions and be more independent, this may actually cause a spike in interactions with the instructor. Indeed, another project analyzing the interactions between student groups in this dataset found that students interact less with their peers during the project lab than in previous labs [28]. We hypothesize that students may seek more help from the instructor when given more agency in the classroom such that different groups are investigating different topics.

Another distinction between the lab types was the student population. The traditional and old reformed labs were part of a course comprised primarily of physics majors, while the new reformed labs were targeted for engineering and other science majors. The diversity of these populations (in terms of race or ethnicity, gender, and academic level) also differed significantly. If the diversity of the sections was a significant influence on instructor-student interactions, we would expect to see more similarities between the traditional section and the other sections.

Additionally, variability between sessions may be due to external events (such as course examinations or upcoming holidays) or myriad other factors. These and other hypotheses should be explored by analyzing the activities themselves in more detail, such as through content analysis on the lab instructions for levels of inquiry [31] or support for student agency [8,36], or other activities taking place in the course or institution.

V. CONCLUSIONS AND FUTURE WORK

This work is part of a broader research program to evaluate the degree to which student-instructor interactions may be a responsible mechanism for the improved outcomes observed from nontraditional labs [3]. In this study, we evaluate how the degree of interaction between instructors and students varies between different lab types. We find that sections of the nontraditional, reformed labs show, on average, higher overall student-instructor interactions than the traditional labs.

The work here also suggests that interactions between instructors and students may be more a product of the lab design than an individual instructor's teaching choices.

All instructors in this dataset, however, were experienced instructors, were willing to have us record their teaching, and, in the reformed labs, supported the goals of the reformed labs. Novice instructors often report feeling inadequate in teaching roles or that they lack the necessary training to facilitate interactions with students in the lab setting [37,38], despite enthusiasm to implement evidence-based practices [38]. Any conclusions about traditional labs are particularly limited given that we only had data from one traditional lab section. Future work should further replicate this study with more diverse and novice instructors to test the claims put forth.

Another significant limitation is that this work was conducted at one institution with only one type of reformed physics lab. Future work should apply our methodologies at other institutions and with broader populations of students. We are currently evaluating whether computer vision techniques can extract information about the interactions, as the methods here do not require one to evaluate the content of the interactions. If successful, computer vision techniques would facilitate a large scale study of lab instruction from video data at multiple institutions.

Future work should also seek to evaluate what parts of the reformed lab activities best facilitated interactions between students and instructors. That is, what makes a "good" lab in terms of optimizing student-instructor interactions? Are more interactions necessarily better or can more interactions be a result of an overly difficult or confusing assignment?

Based on the confounding variables and correlational nature of the study, we are hesitant to suggest implications for instruction. We tentatively put forth that the data support having regular training meetings with instructors that focus on facilitating interactions (as implemented in the new reformed labs but not the other two lab types).

ACKNOWLEDGMENTS

We are greatly appreciative of the feedback from Eric Brewe, Michelle Smith, and David Esparza on this work, from Daniyar Kushaliev in coding some of the lab video, and from Emily Smith, Anna Phillips, and Scott Allen for coordinating the video data collection. This work was supported in part by the Cornell University Active Learning Initiative.

---

[1] Bethany R. Wilcox and H. J. Lewandowski, Open-ended versus guided laboratory activities: Impact on students' beliefs about experimental physics, Phys. Rev. Phys. Educ. Res. **12**, 020132 (2016).

[2] Bethany R. Wilcox and H. J. Lewandowski, Developing skills versus reinforcing concepts in physics labs: Insight from a survey of students' beliefs about experimental physics, Phys. Rev. Phys. Educ. Res. **13**, 010108 (2017).

[3] Emily M. Smith, Martin M. Stein, Cole Walsh, and N. G. Holmes, Direct Measurement of the Impact of Teaching Experimentation in Physics Labs, Phys. Rev. X **10**, 011029 (2020).

[4] Eugenia Etkina, Alan Van Heuvelen, Suzanne White-Brahmia, David T. Brookes, Michael Gentile, Sahana Murthy, David Rosengrant, and Aaron Warren, Scientific abilities and their assessment, Phys. Rev. ST Phys. Educ. Res. **2**, 020103 (2006).

[5] Eugenia Etkina, Anna Karelina, Maria Ruibal-Villasenor, David Rosengrant, Rebecca Jordan, and Cindy E. Hmelo-Silver, J. Learn. Sci. **19**, 54 (2010).

[6] Anna Karelina and Eugenia Etkina, Acting like a physicist: Student approach study to experimental design, Phys. Rev. ST Phys. Educ. Res. **3**, 020106 (2007).

[7] N. G. Holmes, Carl E. Wieman, and D. A. Bonn, Teaching critical thinking, Proc. Natl. Acad. Sci. U.S.A. **112**, 11199 (2015).

[8] N. G. Holmes, Benjamin Keep, and Carl E. Wieman, Developing scientific decision making by structuring and supporting student agency, Phys. Rev. Phys. Educ. Res. **16**, 010109 (2020).

[9] Tong Wan, Constance M. Doty, Ashley A. Geraets, Christopher A. Nix, Erin K. H. Saitta, and Jacquelyn J. Chini, Evaluating the impact of a classroom simulator training on graduate teaching assistants' instructional practices and undergraduate student learning, Phys. Rev. Phys. Educ. Res. **17**, 010146 (2021).

[10] Jared B. Stang and Ido Roll, Interactions between teaching assistants and students boost engagement in physics labs, Phys. Rev. ST Phys. Educ. Res. **10**, 020117 (2014).

[11] Tong Wan, Ashley A. Geraets, Constance M. Doty, Erin K. H. Saitta, and Jacquelyn J. Chini, Characterizing science graduate teaching assistants' instructional practices in reformed laboratories and tutorials, Int. J. STEM Educ. **7**, 12 (2020).

[12] Katherine L. McNeill, Lisa M. Marco-Bujosa, María González-Howard, and Suzanna Loper, Teachers' enactments of curriculum: Fidelity to Procedure versus Fidelity to Goal for scientific argumentation, Int. J. Sci. Educ. **40**, 1455 (2018).

[13] Matthew Wilcox, Yuehai Yang, and Jacquelyn J. Chini, Quicker method for assessing influences on teaching assistant buy-in and practices in reformed courses, Phys. Rev. Phys. Educ. Res. **12**, 020123 (2016).

[14] Benjamin T. Spike and Noah D. Finkelstein, Design and application of a framework for examining the beliefs and practices of physics teaching assistants, Phys. Rev. Phys. Educ. Res. **12**, 010114 (2016).

[15] Renee Michelle Goertzen, Rachel E. Scherr, and Andrew Elby, Tutorial teaching assistants in the classroom: Similar teaching behaviors are supported by varied beliefs about teaching and learning, Phys. Rev. ST Phys. Educ. Res. **6**, 010105 (2010).

[16] Melissa Dancy, Charles Henderson, and Chandra Turpen, How faculty learn about and implement research-based instructional strategies: The case of Peer Instruction, Phys. Rev. Phys. Educ. Res. **12**, 010110 (2016).

[17] Jonathan B. Velasco, Adam Knedeisen, Dihua Xue, Trisha L. Vickrey, Marytza Abebe, and Marilyne Stains, Characterizing instructional practices in the laboratory: The laboratory observation protocol for undergraduate STEM, J. Chem. Educ. **93**, 1191 (2016).

[18] Cassandra Paul and Emily West, Using the real-time instructor observing tool (RIOT) for reflection on teaching practice, Phys. Teach. **56**, 139 (2018).

[19] David Esparza, Amy E. Wagler, and Jeffrey T. Olimpo, Characterization of instructor and student behaviors in CURE and non-CURE learning environments: Impacts on student motivation, science identity development, and perceptions of the laboratory experience, CBE Life Sci. Educ. **19**, ar10 (2020).

[20] Danny Doucette, Russell Clark, and Chandralekha Singh, All aboard! Challenges and successes in professional development for physics lab TAs, *Proceedings of the 2019 Physics Education Research Conference, Provo, UT*, (AIP, New York, 2020).

[21] Danny Doucette, Russell Clark, and Chandralekha Singh, Professional development combining cognitive apprenticeship and expectancy-value theories improves lab teaching assistants' instructional views and practices, Phys. Rev. Phys. Educ. Res. **16**, 020102 (2020).

[22] Jacquelyn J. Chini, Carrie L. Straub, and Kevin H. Thomas, Learning from avatars: Learning assistants practice physics pedagogy in a classroom simulator, Phys. Rev. Phys. Educ. Res. **12**, 010117 (2016).

[23] Luke Lefebvre and Mike Allen, Teacher immediacy and student learning: An examination of lecture/laboratory and self-contained course sections, J. Scholarship Teach. Learn. **14**, 29 (2014).

[24] N. G. Holmes and Emily M. Smith, Operationalizing the AAPT learning goals for the lab, Phys. Teach. **57**, 296 (2019).

[25] Z. Yasemin Kalender, Emily Stump, Katelynn Hubenig, and N. G. Holmes, Restructuring physics labs to cultivate sense of student agency, Phys. Rev. Phys. Educ. Res. **17**, 020128 (2021).

[26] Olivier Friard and Marco Gamba, BORIS: A free, versatile open-source event-logging software for video/audio coding and live observations, Methods Ecol. Evol. **7**, 1325 (2016).

[27] Cole Walsh, Daniyar Kushaliev, and N. G. Holmes, Connecting the dots: Student social networks in introductory physics labs in *Proceedings of the 2020 Physics Education Research Conference*, Virtual Conference, edited by Steven Wolf, Michael B. Bennett, and Brian W. Frank (AIP, New York, 2020), pp. 557–562.

[28] Meagan Sundstrom, David G. Wu, Cole Walsh, Ashley B. Heim, and N. G. Holmes, Examining the effects of lab instruction and gender composition on intergroup interaction networks in introductory physics labs, Phys. Rev. Phys. Educ. Res. **18**, 010102 (2022).

[29] Ulrik Brandes and Thomas Erlebach, Fundamentals, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, edited by Ulrik Brandes and Thomas Erlebach (Springer, Berlin, Heidelberg, 2005), Vol. 3418, pp. 7–15.

[30] Nicholas Bogel-Burroughs, Deer Dies in Cornell Graduate Student Office After Leaping Through Rockefeller Hall Window, The Cornell Daily Sun, 2017, https://cornellsun.com/2017/11/16/deer-dies-in-cornell-graduate-student-office-after-leaping-through-rockefeller-hall-window/.

[31] Laura B. Buck, Marcy H. Towns, and Stacey Lowery Bretz, Research and teaching: Characterizing the level of inquiry in the undergraduate laboratory., J. Coll. Sci. Teach. **38**, 52 (2008), https://eric.ed.gov/?id=EJ809323.

[32] This activity was modeled on one developed by David Marasco.

[33] Anna McLean Phillips, Meagan Sundstrom, David G. Wu, and N. G. Holmes, Not engaging with problems in the lab: Students' navigation of conflicting data and models, Phys. Rev. Phys. Educ. Res. **17**, 020112 (2021).

[34] Ashley A. Geraets, Isadore L. Nottolini, Constance M. Doty, Tong Wan, Jacquelyn J. Chini, and Erin K. H. Saitta, Preparing GTAs for active learning in the general chemistry lab: Development of an evidence-based rehearsal module for a mixed-reality teaching simulator, J. Sci. Educ. Technol. **30**, 829 (2021).

[35] Emily A. West, Cassandra A. Paul, David Webb, and Wendell H. Potter, Variation of instructor-student interactions in an introductory interactive physics course, Phys. Rev. ST Phys. Educ. Res. **9**, 010109 (2013).

[36] Emily M. Smith and N. G. Holmes, Evaluating instructional labs' use of deliberate practice to teach critical thinking skills, Phys. Rev. Phys. Educ. Res. **16**, 020150 (2020).

[37] Ashley B. Heim and Emily A. Holt, Benefits and challenges of instructing introductory biology course-based undergraduate research experiences (CUREs) as perceived by graduate teaching assistants., CBE Life Sci. Educ. **18**, ar43 (2019).

[38] Emma C. Goodwin, Jane N. Cao, Miles Fletcher, Justin L. Flaiban, and Erin E. Shortlidge, Catching the wave: Are biology graduate students on board with evidence-based teaching?, CBE Life Sci. Educ. **17**, ar43 (2018).