Problematizing in inquiry-based labs: how students respond to unexpected results

Meagan Sundstrom, Anna McLean Phillips, and N. G. Holmes
Laboratory of Atomic and Solid State Physics, Cornell University, 245 East Avenue, Ithaca, NY, 14853

Problematizing is a physics practice involving the articulation of a gap in understanding into a clear question or problem. Inquiry-based labs may be conducive to problematizing behaviors, as students often collect data that do not agree with simplified models or their intuitive predictions. In this study, we analyzed video of students performing a lab in which they find the acceleration of an object in flight to be different from what the presented models predict. We aimed to identify the various activities that groups engaged in upon recognizing this inconsistency. Common problematizing activities included explicit discussions of physics concepts, proposing a new experiment or calculation, and checking experimental calculations. We found that each group's sequence and duration of activities was quite unique, highlighting a diversity of approaches taken to address this inconsistency.

I. INTRODUCTION

There has been growing emphasis on studying students' engagement in the practices of physics within classrooms [1]. That is, rather than simply learning physics content, students should be supported to engage in activities that resemble and mirror those of professional physicists. Understanding how these practices play out within physics classrooms is an important step in designing curricula that support students' skills within these practices. In this paper, we examine the practice of *problematizing*, the process of refining an uncertainty or inconsistency into a clear problem or question [2, 3]. Problematizing is an important activity in physics [2] that shapes how groups collectively engage in science [4].

Instructional laboratories (labs) provide a natural space where students can grapple with inconsistencies between models, their expectations, and data. Within these labs, problematizing can lead students to independently engage in scientific practices such as refining physics models [5], troubleshooting [6], improving experimental procedures [7], or developing a new research question [8]. Students may stumble upon the inconsistencies or uncertainties naturally [2], they may be prompted to find them [7], or they may be deliberately set up to face them [9, 10]. Some researchers have argued that deliberately structuring labs to create conflict between data and models is productive for helping students understand the tentative nature of physical models [11] or develop skills to refine their procedures and models [9, 10].

Other researchers, however, have found that students may respond to the deliberate conflict unproductively: they may attempt to ignore the problem or engage in questionable research practices [12]. Such questionable research practices are generally associated with attempting to confirm canonical equations [13] and linked to students' overwhelming expectations that the goal of lab activities is to reinforce and confirm phenomena taught in lectures [12, 14]. Students' dismissal of their experiences in favor of canonical models resembles that of students disregarding their intuition in learning physics concepts [15], particularly after instruction that uses an "Elicit-Confront-Resolve" pedagogy [16]. As in other instructional settings, students in labs need to be supported to productively manage the conflict between their expectations, experiences, and formal physics knowledge. That is, they need to be supported in problematizing.

We aim to understand factors that lead students to acknowledge and productively respond to such conflicts and those that lead them to ignore or unproductively respond. As a first step, we examined students problematizing productively in labs, beginning with two research questions:

Q1 What activities do these students engage in when they are confronted with data that conflict with their expectations?
Q2 What patterns exist within and across lab groups?

To answer these questions, we analyzed four groups of students in one lab session and developed a coding scheme to document their different activities when they faced a disagreement between their data and their theoretical models and predictions. In this lab session, students were testing two models for forces on objects in flight: a gravity-only model and a model that included air resistance. Students collected data, however, that reflected a significant impact of buoyancy. We explicitly studied groups that productively problematized around the presence of a buoyant force. Our coding scheme captures the groups' problematizing activities, such as discussing physics concepts and forces, checking their calculations, and proposing new measurements or calculations, among others. Each group engaged in a unique trajectory of activities in response to the problem.

II. METHODS

A. Context

The data for this study come from a lab section for an introductory mechanics course designed for engineering students at a large private research university in the northeastern US. There were 221 students enrolled in the course, three quarters of whom were freshman. The groups analyzed in the present study come from one laboratory section of 19 students, including 8 female students and 11 male students. An experienced teaching assistant (TA) taught the section.

In this course, there are ten lab sessions, with four main experiments. These labs are designed around student agency and experimentation skills rather than reinforcing lecture concepts. Video and audio data of each group allowed us to capture group activities in detail.

These data come from the objects in flight lab, the second main experiment. At the beginning of the lab session, students are asked to predict the acceleration of objects moving vertically using two models: (1) the force of gravity is the only force acting on the object and (2) air drag and the force of gravity are the only forces acting on the object. Students pick an object (beach ball, basketball, coffee filter, etc.) and find the acceleration using a motion sensor. For the balls in particular, there is a measurable effect from the buoyant force, therefore setting up a conflict between students' data and the predictions they make for when the ball is moving upwards.

B. Episode selection

There were seven groups of 2-3 students in the lab. Each group's video was watched in its entirety for the 2-hour lab session, with notes summarizing the group's activities at 5 minute intervals. Four of the seven groups obtained accelerations smaller in magnitude than 9.8 m/s^2 *and* directly contrasted this to their expectations that the acceleration of a rising object should be equal to or greater than 9.8 m/s^2 due to the combined impacts of gravity and air drag. A fifth group expressed surprise that their acceleration values were constant and less than 9.8 m/s^2; however, much of the conversation that followed was not clearly captured by the audio recorder. For a sixth group, the audio quality was poor enough throughout the session that we could not determine if they compared their data to the models. The final group did not compare

their results to their models and thus did not problematize. This gave us four episodes to analyze.

For the four groups that did problematize and were clearly recorded, we noted when they made a clear statement about the inconsistency between their data and their predictions. This statement was taken as the start of the episode for analysis. Two episodes concluded when the teaching assistant began a class discussion, and the other two concluded when the students began a new experiment by testing a new falling object. Because problematizing is the activity of developing a question, moving on to answer the question can conclude an episode of problematizing. These episodes were transcribed.

C. Coding scheme development

We developed an initial coding scheme to describe students' communication events [17] while problematizing during this lab session. The codes were refined using the selected episodes, which focused on a single, common uncertainty. In order to capture the nature of these activities, we noted who participated in the discussion (Dimension 1 of the coding scheme), who initiated any interaction outside of the immediate lab table (Dimension 2), and a description of the activity and discussion (Dimension 3). Our codes for Dimension 3 were emergent and were not hypothesized apriori. Coding for all three of these dimensions allowed us to fully characterize students' responses to the problem by capturing who they chose to interact with in addition to the substance of the interaction. Dimension 2 does not tend to the aims of this preliminary study, so we do not present or discuss results for those codes.

Sundstrom and Phillips independently applied the draft coding scheme to each episode, then refined the coding scheme based on discrepancies and edge cases. Holmes then coded all four tables, and Sundstrom and Phillips' consensus codes were compared to Holmes' codes. The inter-rater agreement at this stage was 92% for Dimension 1, 82% for Dimension 2, and 69% for Dimension 3. Substantial disagreements in Group 6 significantly reduced the inter-rater agreement and are discussed below. Inter-rater agreement for Dimension 3 is 75% if Group 6 is excluded. All three researchers discussed code disagreements and together came to an agreement on the final codes.

III. RESULTS

Dimension 3 of our coding scheme represents the activities we identified within the problematizing episodes (Research Q1). These codes (Table I) describe the main topic of the interaction, which tells us the activities that students engaged in when their data were in conflict with their expectations. To compare each group's coded problematizing activities (Research Q2), we first describe each group's main activities during the episodes. Fig. 1 summarizes the coded activities for each episode.

TABLE I. Descriptions of Dimension 3 codes.

| Code	Description
Check Calculations	Students check that calculations have been done correctly, including whether or not they have the right data for their calculations. They may explicitly discuss how they got particular numbers or make a statement that they are going to redo calculations.
Check Data	Students check that there are no obvious flaws in the data, including a sensor not measuring accurately. They may look for spikes or suspiciously flat parts of the data that indicate errors in measurement.
Physical Reasoning	Students are using concepts of physics (e.g. forces) to describe either their expectations for the outcome or to make sense of their results. They may also throw objects in the air, use hand gestures, draw and/or point towards the lab equipment.
Report Writing	The primary activity the students are engaging in is discussing what should be written in the lab report. Note that they may include other statements in this discussion (stating reasoning previously discussed, restating their results), but they are not describing new reasoning or new results.
Consult Reference	Students consult a reference, such as the lab manual or online resources.
State Results	Students state their experimental results without any of the above codes applying to the same event.
Propose New	Students propose (but do not execute) new experiments, measurements, or different types of calculations. Students are suggesting that they perform a new activity to gain a better understanding or additional information.
Change Subject	Students change the subject away from the problem and toward something else related to the lab activity, for example proposing to repeat the experiment with a different object without a discussion of why.
Off Topic	This code applies to conversations not relevant to the lab, such as telling a joke or discussing weekend plans, hobbies, or unrelated coursework.
Silence	This code applies to silence of more than 5 seconds during which students do not appear to be performing a task related to the lab (ex: performing additional data collection or writing lab notes).
Inaudible	Enough speech is inaudible that it is not possible to apply one of the above codes.
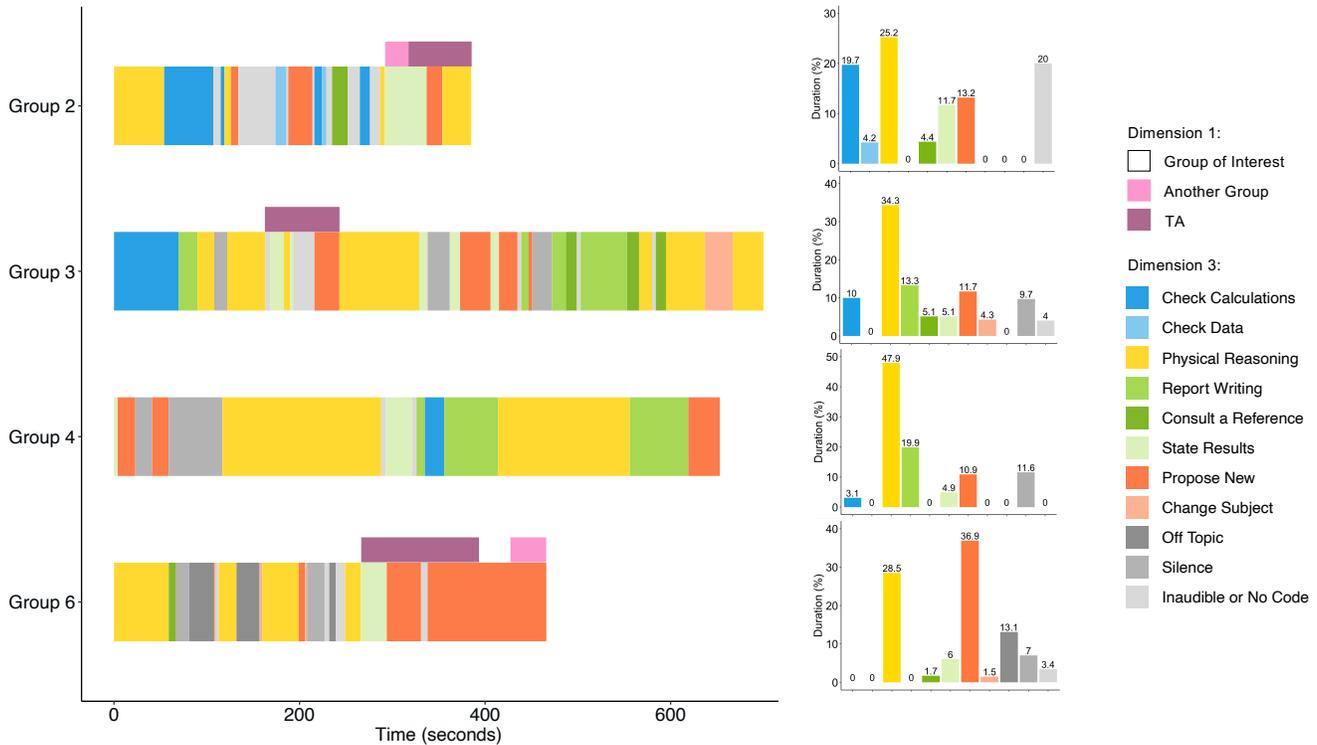
FIG. 1. The scarf plots on the left illustrate the finalized sequence of Dimension 1 (top bar) and Dimension 3 (lower bar) codes applied to each group's episode. The bar charts on the right show the percentage of the problematizing episode spent within each Dimension 3 activity.

Group 2 consisted of two female students and one male student, and this was the shortest episode we analyzed. They begin by debating the impact of drag on the acceleration of a falling object. They then check their data and calculations, highlight a different portion of data, and recompute the acceleration. Later, they discuss how to analyze separate bounces of the object and decide to ask the TA for help. This episode ends when the TA extends a comment to this group to an entire class discussion.

Group 3 consisted of one female student and one male student, and this was the longest episode we analyzed. After checking their calculations, they determine that there must be some upward force acting on the ball. They consult the lab manual for directions, and use Google to look up how the force of lift acts on an airplane. At the end of the episode, they try to relate this lab to a homework problem, and the episode ends when the TA begins the entire class discussion that also ends the episode for Group 2.

Group 4 consisted of three male students. They do not interact with the TA or another group and they spend a notable amount of time in silence or writing their lab notes. They begin by brainstorming a hypothesis to falsify, and they try to come up with upward forces that could be acting on the falling object. One group member suggests buoyancy, but the idea is dropped. They decide to test "air flow," and the episode ends when they collect data with a different object.

Group 6 consisted of two male students and one female

student, and distinctly spends the most time off-topic and apparently joking. They realize that there must be an upward force acting on the ball, and their main idea of what the force could be – the moon pulling the ball upwards – is brought up consistently throughout the episode. A group member jokes about napping on the lab table and suggests that may impact the object's acceleration. They initiate a conversation with the TA about what to do next, and the episode ends when they decide to test a different object. Whether or not to code these joking moments as "Off Topic" or another code was a significant challenge for the coders, and the inter-rater reliability for this group was very low in Dimension 3 (53%). We ultimately decided to code their statements based on the content of their speech (such as coding the moon statements as "Physical Reasoning") because, although joking, the speech was not off-topic.

Figure 1 allows us to visually compare each group's sequence and duration of problematizing behaviors. The scarf plots on the left illustrate our finalized consensus codes applied to each group for the duration of the episode, while the bar charts on the right display the percentage of total episode time spent in each code within Dimension 3.

IV. DISCUSSION AND CONCLUSION

In this study, we developed a coding scheme to identify and evaluate students' activities when problematizing during a physics lab. Though we analyzed four groups, five of the

seven groups were identified as engaging in problematizing. Because one of these groups was inaudible for most of of the session, we only know that one group did not problematize at all. This suggests that purposely designing uncertainties or inconsistencies in a lab can prompt problematizing.

The coding scheme summarizes what students do in response to puzzling experimental results (Table I), and with whom they have these interactions. Problematizing behaviors included physical reasoning, checking calculations, checking data, report writing, off-topic conversation, stating results, and proposing new experiments or calculations. Most groups engaged in a variety of activities, and most activities were identified in more than one group.

Across the episodes we analyzed, we see that physical reasoning was a dominant activity. This makes sense: when these students were confronted with this uncertainty, they used resources from their conceptual understandings to refine and make sense of what they do not know. Significant engagement in physical reasoning is in line with the purpose of inquiry-based physics labs; we want students to think deeply about their experiments and push simplified physical models to their limits. Another recurring activity for all of these groups was proposing new experiments or calculations. This is expected in the inquiry-based lab setting, as students develop and iterate their own experimental procedures.

Three of the four analyzed groups interacted with the TA within the episode, and all of these discussions looked similar: the groups stated their results, engaged in physical reasoning, and proposed new ways forward in their experiment. This pattern may reflect the TA's moves, students' questions for the TA, or a combination thereof and merits further study.

The bar charts in Figure 1 convey that while the most dominant activities across the groups are fairly similar, the distribution of codes exhibited by each group is quite different. The scarf plots in Figure 1 highlight the diversity among these four groups in how they wrestle with the problem, as each went through various amounts, lengths, and sequences of behaviors. Though these groups were sitting only a few feet away from each other and were supported by the same TA and lab materials, they all took different approaches to dealing with the same inconsistency. After verbally acknowledging an inconsistency between their experimental results and the given physical models, the groups engaged in different initial activities: Two groups (2 and 6) began with physical reasoning, one group (3) went back and checked their calculations, and one group (4) briefly stated their results and then proposed a new idea. Despite the common patterns within TA-group interactions, the uniqueness of each group's approach to the issue suggests that instructors may need multiple strategies when supporting students' problematizing.

One major difference across these four groups is that Group 4 switches between activities at a much lower frequency than the other three groups (1.1 switches per minute versus more than two switches per minute). This reflects the slower pace of discussion. They are also the only group that did not speak to the TA or a member of another group. Fu-

ture work should examine whether this activity switching is indicative of the quality of students' problematizing or personalities in the group.

Group 6 was the only group that did not check data or calculations during the episode; they also are the only group to engage in off topic conversation, which they did for a significant amount of time. Group 6's timeline suggests that some students may be reluctant or uncomfortable in facing uncertainties and engaging in problematizing. This suggests that intellectual humility [18] or meta-affective learning [19] may be relevant lenses for understanding students' problematizing. While an intellectually humble individual effectively acknowledges and responds to their knowledge gaps [18], some students may find this admittance discomforting [19, 20]. Changing the subject of the conversation or joking about other topics allows them to deviate attention away from the issue [20]. At the same time, it is possible to shift students' responses to encountering uncertainty in the classroom [19]. Instructors wishing to foster problematizing should create a learning environment in which students are comfortable with being puzzled during group work.

The variety of activities seen in our data suggests that dealing with an uncertainty in a lab allows students to gain valuable scientific skills: thinking critically about their experimental design and responding to a confusing or unexpected situation [2]. These results contrast with other research that indicates that students may respond unproductively when an experiment or demonstration shows an unexpected result [12, 15]. Further understanding of the dynamics of both productive problematizing and these unproductive responses is necessary for navigating the complex instructional space where students' expectations, prior knowledge, or intuition are apparently inconsistent with real-world phenomena.

Future work will examine additional groups to further validate the coding scheme and identify any behaviors that may arise in different scenarios. A broader analysis will aid in identifying additional patterns, including patterns related to who initiated interactions (Dimension 2 of the coding scheme, which was omitted from this analysis). We are also studying the group that did not problematize in this lab session to understand how their activities and engagement differ from the groups analyzed here. We focused on students' activities within a bounded problematizing episode, but future work may apply the coding scheme to the entirety of the lab section to examine whether these activities occur more broadly. Finally, studying how the instructor's moves influenced students' activities can provide insight into how instructors can support students' problematizing.

[1] National Research Council, *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering* (National Academies Press, 2012).

[2] A. M. Phillips, J. Watkins, and D. Hammer, Problematizing as a scientific endeavor, Physical Review Physics Education Research **13**, 020107 (2017).

[3] A. M. Phillips, J. Watkins, and D. Hammer, Beyond "asking questions": Problematizing as a disciplinary activity, Journal of Research in Science Teaching **55**, 982 (2018).

[4] J. Watkins, D. Hammer, J. Radoff, L. Z. Jaber, and A. M. Phillips, Positioning as not-understanding: The value of showing uncertainty for engaging in science, Journal of Research in Science Teaching **55**, 573 (2018).

[5] B. M. Zwickl, D. Hu, N. Finkelstein, and H. J. Lewandowski, Model-based reasoning in the physics laboratory: Framework and initial results, Physical Review Special Topics - Physics Education Research **11**, 020113 (2015).

[6] D. R. Dounas-Frazer, K. L. Van De Bogart, M. R. Stetzer, and H. J. Lewandowski, Investigating the role of model-based reasoning while troubleshooting an electric circuit, Physical Review Physics Education Research **12**, 010137 (2016).

[7] N. Holmes, C. E. Wieman, and D. Bonn, Teaching critical thinking, PNAS **112**, 11199 (2015).

[8] E. Etkina, A. Karelina, M. Ruibal-Villasenor, D. Rosengrant, R. Jordan, and C. E. Hmelo-Silver, Design and Reflection Help Students Develop Scientific Abilities: Learning in Introductory Physics Laboratories, Journal of the Learning Sciences **19**, 54 (2010).

[9] P. A. Bartlett and K. Dunnett, Secret objectives: promoting inquiry and tackling preconceptions in teaching laboratories, (2019), arXiv:1905.07267.

[10] M. Vonk, P. Bohacek, C. Militello, and E. Iverson, Developing model-making and model-breaking skills using direct measurement video-based activities, Physical Review Physics Education Research **13**, 020106 (2017).

[11] E. Brewe, Modeling theory applied: Modeling Instruction in introductory physics, American Journal of Physics **76**, 1155 (2008).

[12] E. M. Smith, M. M. Stein, and N. Holmes, How expectations of confirmation influence students' experimentation decisions in introductory labs, Physical Review Physics Education Research **16**, 010113 (2020).

[13] M. M. Stein, E. M. Smith, and N. Holmes, Questionable research practices in introductory physics labs, Bulletin of the American Physical Society (2020).

[14] D. Hu and B. M. Zwickl, Examining students' personal epistemology: the role of physics experiments and relation with theory, in *2017 Physics Education Research Conference Proceedings* (American Association of Physics Teachers, 2018) pp. 11–14.

[15] E. F. Redish and D. Hammer, Reinventing college physics for biologists: Explicating an epistemological curriculum, American Journal of Physics **77**, 629 (2009).

[16] L. C. McDermott, Millikan Lecture 1990: What we teach and what is learned-Closing the gap, American Journal of Physics **59**, 301 (1991).

[17] S. Hennessy, S. Rojas-Drummond, R. Higham, A. M. Márquez, F. Maine, R. M. Ríos, R. García-Carrión, O. Torreblanca, and M. J. Barrera, Developing a coding scheme for analysing classroom dialogue across educational contexts, Learning, Culture and Social Interaction **9**, 16 (2016).

[18] D. Whitcomb, H. Battaly, J. Baehr, and D. Howard-Snyder, Intellectual humility: Owning our limitations, Philosophy and Phenomenological Research **94**, 509 (2015).

[19] J. Radoff, L. Z. Jaber, and D. Hammer, "it's scary but it's also exciting": Evidence of meta-affective learning in science, Cognition and Instruction **37**, 73 (2019).

[20] L. D. Conlin and R. E. Scherr, Making Space to Sensemake: Epistemic Distancing in Small Group Physics Discussions, Cognition and Instruction , 1 (2018).