How expectations of confirmation influence students' experimentation decisions in introductory labs

Emily M. Smith[1,2] Martin M. Stein[1] and N. G. Holmes[1]

[1]*Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA*
[2]*Department of Physics, Colorado School of Mines, Golden, Colorado 80401, USA*

Many instructional physics labs are shifting to teach experimentation skills, rather than to demonstrate or confirm canonical physics phenomena. Our previous work found that many students engage in questionable research practices in attempts to confirm the canonical physics phenomena, even when confirmation is explicitly not the goal of the lab. This exploratory study aimed to answer three research questions: (RQ1) What are students' expectations about the purpose of labs when they enter introductory physics?, (RQ2) How do their prior experiences shape those expectations?, (RQ3) In what ways do those expectations relate to their engagement in questionable research practices? Through open-response surveys, we found that students overwhelmingly expressed confirmatory beliefs about the purpose of labs. Through interviews, we found that students' prior lab experiences were also overwhelmingly confirmatory, despite varying degrees of structure. We then used video of individual groups to explore the ways in which questionable research practices manifest through confirmatory expectations. We confirm previous work that students' confirmatory expectations can lead them to engage in questionable research practices, but find that these behaviors occur despite instructional messaging about an alternative purpose. Our analyses also suggest that engagement in questionable research practices is more frequent than the previous results indicated through analysis of submitted lab notes. These results further illuminate issues with traditional labs, but suggest that the confirmatory goals, perhaps more so than high structure, are problematic.

I. INTRODUCTION

For the past 100 years, instructional physics laboratories (hereafter referred to as labs) have been criticized as being inauthentic [1,2]. Most criticisms argue against the highly structured nature—structured such that students can follow procedures without thinking. Recent innovations in labs have focused on removing the structure and providing students with various degrees of freedom over the design and execution of experiments [3–5]. While these instructional innovations address concerns over how students are expected to approach investigations, little attention has been paid to the nature of what students investigate in labs. In this paper, we explore evidence that an emphasis on confirming models in labs may be just as or more problematic than highly structured activities.

Despite many calls to shift the focus of labs to teach students about the nature of science and to develop students' scientific abilities [6–10], introductory physics labs are often used to confirm that the physics content presented in a course holds in the "real world." As such, many introductory students expect labs to supplement their learning of physics content introduced in the lecture parts of the course [11,12] or to confirm results predicted by theory [13]. Little is known, however, about how students' confirmatory expectations interact with instructional goals centered around experimentation, particularly in labs where students are not expected, nor required, to achieve a particular outcome in the lab. As instructional labs shift away from traditional structures, students may struggle to understand their role in the lab.

This exploratory study aims to evaluate students' transition into labs that aim to engage them in the process of scientific investigations: How do students' prior lab experiences influence their expectations about and behaviors during nontraditional instructional labs? We present an analysis of students' expectations about physics labs as they enter their first college introductory physics lab course and find the majority of students enter physics labs expecting to confirm models (similar to previous results, e.g., Refs. [11–13]). We then present evidence from interviews that these expectations develop from students' prior, and perhaps concurrent, lab experiences in high school and college science courses. Finally, we analyze the impacts of

students' expectations about confirming models on their decision making during a lab designed to engage students in expertlike experimentation practices. We conclude that students' tendency to expect that physics labs are activities in model confirmation leads to investigations aiming to confirm a model, causing them to engage in questionable research practices.

II. THEORETICAL GROUNDING

In this section, we define terms that are relevant to the phenomena we observed in our data. Because of the exploratory nature of this study, our theoretical grounding does not provide a predictive model, and instead, provides context to our methods, data, and analyses as well as informs our choices in vocabulary.

In this paper, we define a *model* to be a representation or coordinated set of representations that describe or predict the behavior of a physical system [14,15]. We describe common expectations that students have when they enter introductory physics labs in college and broadly refer to expectations about confirming models as model confirming expectations. Finally, we define questionable research practices in the context of experimentation to be the decisions or behaviors that call into question the objectivity of experimental results.

A. Role of models in physics labs

The modeling framework for experimental physics explicates the processes involved with model building and refinement through experimental physics [16]. The framework outlines two distinct branches of modeling that expert experimental physicists consider throughout their investigations: modeling of physical and measurement systems.

Modeling of physical systems requires one to synthesize known and relevant physical principles in the context of the physical system. A model of the physical system ("physical model") may be communicated through coordinated representations that include graphs, equations, pictures, and descriptions. Limitations, assumptions, and approximations of the model are identified, and the model should be predictive within this scope. A physical model is refined through iterative, empirical tests. When limitations, approximations, or assumptions of the model become relevant to the physical system, refinements to the model must be made to appropriately describe and predict the behavior of the physical system. Modeling of measurement systems focuses on developing a description of the limitations of a measurement system, which is comprised of measurement tools and techniques. Depending on the required precision of a measurement, the measurement model may be more limited than a physical model. Expert experimental physicists engage in modeling of the physical and measurement systems [17], and lab instructors identify modeling as a desired learning outcome for instructional physics labs [18].

In this paper, we focus on students' expectations about models of physical systems. Models of measurement systems are important in experimental physics, however, our data suggest that introductory students' expectations about measurement models are less developed and robust than their expectations about physical models and that their expectations about lab activities are heavily tied to physical models. We suspect that students' expectations about models of measurement systems may affect their expectations about physical models, however, we chose not to probe these possible relationships in this study.

B. Expectations about labs and experimental physics

Expectations are the set of understandings that a student brings to a course that include "attitudes, beliefs, and assumptions about what sorts of things they will learn, what skills will be required, and what they will be expected to do" [19]. Research has repeatedly shown that students' understanding of and expectations about physics tasks influence the ways in which they engage in a course (e.g., Refs. [20–24]). In this paper, we present evidence that suggests students' engagement in questionable research practices in introductory physics labs occurs due to students' expectations that lab instruction is about confirming models.

We define a *model confirming expectation* to be when a student understands the purpose of a lab to be confirming, verifying, or demonstrating that a physical model provided to them (through lecture, lab materials, textbook, etc.) holds true in the lab. Our definition of a model confirming expectation is broad and described at a large grain size. However, the data suggest that it manifests in distinct ways among individuals and groups.

In this study, we explicitly distinguish model *confirming* from model *testing*. We define model testing to be evaluating whether a physical model may or may not hold for a particular physical situation (whether or not a specific model is provided). The degree to which a model is appropriate to the situation may depend on factors such as the experimental regime or the precision of the measurement. Model testing encompasses everything from selecting among multiple competing models to comparing the outcomes of different experimental or theoretically-driven techniques to probing a model at higher precision. We expect that many physics experts engage in model testing during their experiments, such as designing LIGO to detect gravitational waves (requiring sufficient precision within the regime of interest while acknowledging a, perhaps small, possibility of failure). Whether or not a predictive model exists, model testing maintains ambiguity as to the intended outcome. Science does not aim to obtain a particular result; science seeks truth, whatever that truth may be. In contrast, we define model confirming to be conducting an experiment with the intent to obtain a particular outcome that confirms, verifies, or demonstrates

the theoretical model. Model confirming expectations, therefore, are such that the student: expects that a model provided to them will hold true in the lab, does not expect the activity to offer ambiguity about whether or when the model does or does not apply, and/or does not expect the data to disagree with the model.

As outlined earlier in the introduction, research has documented that students expect their physics labs to be exercises in confirming models and to supplement their learning of course material [11–13]. Some students view physics experiments with the primary purpose of gaining understanding of the course material, not necessarily an explicit activity in model confirmation [11]. Furthermore, model confirming expectations manifest such that many students believe that the validity of results should be judged based on their agreement with a physical model [25]. Less frequently, students mention experimentation for the purpose of discovery [11].

C. Questionable research practices

In scientific research, an explicit aim to verify a particular result—without informed skepticism—may be considered unethical and history demonstrates that these aims have delayed scientific progress [26]. United States federal policy outlines research misconduct to include fabrication, falsification, and plagiarism [27]. Examples of research misconduct, such as making up data or publishing other researchers' works, are blatantly unethical practices in research. However, there is an ethical grey area that includes research practices that are not flagrant cases of research misconduct but may call into question the objectivity of reported results and conclusions. Research practices that "are actions that violate traditional values of the research enterprise and that may be detrimental to the research process" fall into this grey area of questionable research practices [28]. In the literature, questionable research practices range from outright misconduct (e.g., fabricating data) to, perhaps, acceptable norms within a field (e.g., failing to report all models that were tested) [29–32]. Popular examples of questionable research practices include p hacking, cherry picking, and hypothesizing after results are known (HARK-ing), however, no complete list of questionable research practices exists. Researchers who engage in research misconduct usually do so with intention, however, those who engage in questionable research practices may have no intention to mislead and may operate within disciplinary norms. In this paper, we will use the term questionable research practices to mean decisions or behaviors that call into question the objectivity of experimental results.

Previous research suggests that some introductory physics students engage in questionable research practices, especially interpreting or manipulating data to obtain, unjustifiably so, particular results. For example, students have exhibited difficulties coordinating claims and evidence from complex datasets and made claims based on their prior knowledge rather than the data available to them [33]. In lab activities, some students concealed systematic errors by inflating experimental uncertainties to obtain agreement between data and a model [34–36]. These results hint that questionable research practices may be common in students' analyses and interpretations of data, which are central skills in experimentation labs. Other evidence suggests that students may develop justification that supports strong intuitive responses to physics problems [37].

Previously, we conducted a study that emergently identified students' engagement in questionable research practices through lab notes [34]. Most of these questionable research practices were accompanied by language suggesting the students were attempting to confirm the provided model. We also detailed one group's approach during the same investigation [35]. The students explicitly discussed that they assumed the purpose of the activity was to verify the physical model provided in the lab instructions. Their subsequent decisions in the lab were motivated by this shared understanding, which were misaligned with the instructional intent of the activity. The students engaged in various questionable research practices as a result, such as disregarding instructional hints and intentionally inflating their experimental uncertainties by reverting to methods that yielded less precise measurements.

These findings drove the analyses presented in this paper; here, we aim to develop testable explanations for students' expectations and their resulting engagement in questionable research practices. In choosing to follow this line of research, we deliberately selected data and analyses that speak to students' expectations and questionable research practices. We intend to shed light on the consequences of confirmatory expectations in labs that are designed to engage students in nonconfirmatory activities.

D. Research questions

Our research questions are as follows:
- What are students' expectations about labs when they first enter their introductory physics labs? This question serves to test whether the beliefs found in previous work [11–13] also existed in our sample.
- Do students' model confirming expectations about labs come from their other lab experiences? Prior work has not explored the reasons why students enter their physics labs expecting that the lab materials will help them learn the course content. To explore possible reasons why students may have these expectations, we interviewed students about their prior and concurrent lab experiences.
- In what ways do students' model confirming expectations about labs influence engagement in questionable research practices? The instructional goals of the lab activity described in this paper were explicitly not centered around model confirmation, which directly

conflicts with many students' expectations. We iden-
tified questionable research practices that students
engaged in when they set out to confirm a model,
extending our previous work.

III. METHODS

This study is an exploratory study into students' engage-
ment in questionable research practices due to their model
confirming expectations about physics labs. We previously
identified questionable research practices and model con-
firming expectations in groups' lab notes [34] and sought
data and analyses to identify the origins and implications of
these practices. However, we present the results in an order
different from our process of data analysis. Figure 1
provides information for the order of the analysis process
and how different data sources were motivated and col-
lected based on prior observations.

The majority of data in this study are from a mechanics
course that is aimed at prospective physics majors and
minors ("Physics"). Most students who choose to enroll in
the course are highly prepared for introductory mechanics,
having taken a college-level physics course in high school
(e.g., AP Physics). All students in this course were expected
to be concurrently enrolled in a multivariable calculus
course. Typical demographics for this course include about
20% women and 80% men, 10% students from under-
represented racial or ethnic groups in physics, and 50%
students who intend to pursue physics majors [38].

A portion of data are from the engineering course for
introductory mechanics that is aimed at prospective engi-
neering, science, and mathematics majors ("Engineering").
Typical demographics for this course include about 55%
women and 45% men, 15% students from underrepresented
racial or ethnic groups in physics, 5% students who intend
to pursue physics majors, 85% who intend to pursue
engineering majors, and 10% who intend to pursue other
science (nonphysics) or mathematics majors.

A. Instructional context

The instructional labs in this study were designed to
teach students about the nature of experimentation and
develop various critical thinking skills and scientific
practices by requiring students to make decisions about
the design and execution of their experiments and regularly
extending experiments (see Refs. [5,39] for additional
descriptions). Structure was faded from the lab activities
as the semester progressed, which left more decisions about
the investigations to the students. The activity in this paper
was the first of the semester, and so contained the most
structure. By the final activity of the semester, groups
developed a research question to extend an investigation
from one of the earlier activities, and so, nearly all decisions
were made by the students.

Lab note analysis: An analysis of groups' lab notes revealed many instances of decisions that appeared to be driven by students' desire to confirm the model provided to them in the lab instructions. This initial, unexpected finding motivated the project. Presented in Ref. [34].

Video analysis (preliminary): We analyzed several videos to investigate how groups were approaching the lab activity (e.g., aiming to confirm the model) and identified several actions they took that followed from their expectations about the activity. Partially presented in Ref. [35].

Interviews: We developed the interview protocol with the intent to probe students' reflections on their deci-sion making during the lab activity and how the activity shaped their perspectives on approaches to experimen-tation. Partially presented in Ref. [35] and this paper.

Written statement on purpose of labs: In an orien-tation activity to introduce features of the electronic lab notebook, we added a statement to test whether students viewed labs as activities in confirmation be-fore receiving any instruction. Presented in this paper.

Video analysis (questionable research practices): We specifically analyzed video for questionable re-search practices and built broader narrative descrip-tions of evidence for students' motivation and deci-sions during the lab period. Presented in this paper.

Lab note analysis (questionable research practices): We analyzed the lab notes of the two groups that were included in the video analysis to investigate whether they included descriptions of their expectations and engagement in questionable research practices in saved versions of their lab notes. Presented in this paper.
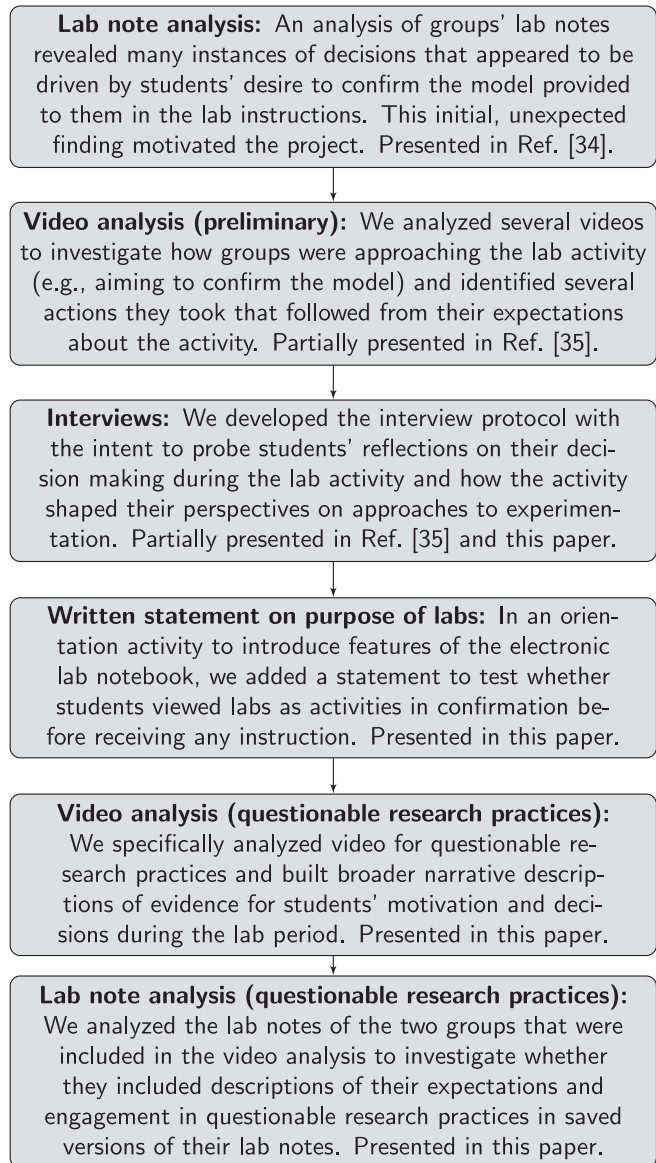
FIG. 1.   Order of data analysis for transparency. The results are
communicated for ease of understanding by the reader and do not
reflect the order of data analysis.

The video and lab notes data presented are from the
activity Pendulum for Pros in the physics course [36]. The
activity was implemented at the beginning of the intro-
ductory lab sequence, and so set the stage for the lab
environment for the semester. Students completed the
activity across two weeks. The instructional goals of the
activity center around iterating to improve measurements
and using statistical tools to evaluate data and inform
experimentation decisions.

The experimental context for the investigation is
testing the angular dependence of the period of a
pendulum. The instructional handout provided a model
for the period:

$$T = 2\pi\sqrt{\frac{L}{g}}.$$

The language (e.g., test, evaluate whether) around this model was deliberate so as to not imply that students were expected to confirm this model through their experiments. The handout prompted students to iterate to improve their measurements (i.e., reduce uncertainty). Instructors emphasized that the labs were designed to engage students in experimentation processes rather than reinforcing physics concepts.

During the first week, through a series of invention tasks [40], students developed a statistic to compare the distinguishability of two measurements within units of uncertainty, referred to as t' [36]. For two measurements, $A \pm \delta A$ and $B \pm \delta B$, t' is expressed as

$$t' = \frac{A - B}{\sqrt{\delta A^2 + \delta B^2}}.$$

The t' value is set up to motivate follow-up decisions in the experiment and students were provided with the following guidelines:

$|t'| < 1$ Indistinguishable measurements, follow-up behaviors should aim to reduce uncertainty

$1 < |t'| < 3$ Inconclusive, follow-up behaviors should aim to reduce uncertainty

$|t'| > 3$ Distinguishable measurements, follow-up behaviors should aim to reduce uncertainty or identify model limitations.

Students then conducted investigations to compare the period of a pendulum when released from 10 and 20 degrees and iterated to improve measurements according to the guidelines. With improved measurements (e.g., additional trials, increasing oscillations per trial), t' increases, and it is possible to experimentally distinguish the periods from 10 and 20 degrees with $t' > 3$, i.e., the small-angle approximation in the model breaks down. With a string length less than 1 m, the periods at these angular displacements are typically distinguishable with several trials consisting of 15 to 20 oscillations per trial or many trials with 5 to 10 oscillations per trial. Many groups also extended their investigation to test other effects (e.g., angular dependence, length of string).

Students worked in groups to conduct their investigations and recorded their experimentation decisions and justifications in an electronic lab notebook. They were instructed to treat the notebook as a journal, with notes addressing the experimental decisions they were making and why these decisions were made. Groups submitted their lab notebooks by the end of the lab period each week.

During the semester that we collected video, the lab component of the course was scored on a pass-fail basis. Students were required to actively participate in 10 out of 11 lab sessions to pass the course. Students received feedback on their lab notes, but the quality of their lab notes did not contribute to their course grade. The lab instructors did not explicitly explain the grading to students, and so, students' expectations were probably tied to their expectations about grading.

B. Identifying students' expectations about labs

At the start of the course, students in both the engineering and physics courses, provided a brief written response to the question "What do you believe is the purpose of a physics lab?" at the beginning of their first lab meeting as part of an introduction to using their electronic lab notebook. Students were prompted to provide brief responses and received credit for submitting the assignment, regardless of their answers. Most lab instructors did not explain the purpose of the labs or the types of lab activities prior to students completing this assignment. Two lab instructors (teaching four sections out of 26), who are both graduate students in the physics education research group, had students familiarize themselves with the overarching learning goals for the labs prior to or during the introduction to the electronic lab notebooks assignment. To respond to these instructional differences, we separated our analyses by the order of the activity (whether students completed the assignment before, during, or after instruction on the goals of the lab).

We emergently coded a subset of about 100 responses; many of the codes resemble those developed by Hu and Zwickl for students' responses to "why are experiments a common part of physics classes?" [12] because we had recently coded other students' responses to that question for another study. Code descriptions are provided in Appendix A. After emergent development of the codes, one rater applied the codes to the remainder of the responses. Then, a second rater separately coded 40 responses and compared code applications between raters. We adjusted codes to clarify definitions and instances for their application. Then, one rater recoded all responses and the second rater coded 50 responses; for these 50 responses, Cohen's kappa was 0.82 and Krippendorff's alpha was 0.84. Note that, due to multiple codes applying to single responses, these reliability statistics are estimates. For individual code comparisons between the two raters, Krippendorff's alpha ranged from 0.65 to 1.0.

Each student's response was characterized by any number of codes. We grouped four of the codes together because they implied a model confirming expectation about the purpose of labs. During discussions, we deliberately chose to apply nonconfirmation codes liberally to capture small references to alternative understandings of the purpose of physics labs. Therefore, the frequencies associated with nonconfirmation codes may be overestimated. When codes were collapsed into the confirmation and nonconfirmation categories Cohen's kappa was 0.88 and Krippendorff's alpha was 0.91.

C. Identifying sources of students' model confirming expectations

We conducted semistructured, one-on-one interviews with 10 students who were enrolled in or had completed the physics course about their expectations and experiences in the lab. In this paper we focus on specific instances when students described their experiences in other lab courses (either college or high school). All students drew some comparisons between this lab and their experiences in other lab activities; in most cases these comparisons were unprompted by the interviewer (see protocol in Appendix B), however, in some cases, the interviewer explicitly asked the student to describe their prior lab experiences.

Five students were interviewed six months after completing the first lab activity, Pendulum for Pros. Five students were interviewed two to three weeks after this first lab. Two interviewers (E. M. S. and M. M. S.) conducted the ten interviews. To maintain consistency between interviewers, we conducted two pilot interviews where one of us observed while the other conducted the interview. After each pilot interview, we compared notes about handling specific situations and modified the interview protocol to include additional information about the timing of follow-up or clarification questioning.

After we completed all of the interviews, N. G. H. listened to the audio recordings of the interviews and marked passages where students reflected on their prior experiences in science labs or compared their experiences in the physics labs to their prior and concurrent lab experiences. We transcribed these portions of the interviews. From the transcripts, we broadly characterized students' descriptions of their other lab experiences and their explanations about how prior experiences influenced their approach to the lab activity or general approaches to labs.

D. Identifying relationships between students' model confirming expectations and engagement in questionable research practices

Lastly, we video-recorded six groups' behaviors and conversations from two lab sections for the second week of the first lab activity in the physics course. Students selected their own groups and locations in the classroom including whether or not to sit at a video-recorded table, so the groups were not formed for specific research purposes. To identify students' expectations about lab activities, we analyzed video by identifying, in all six groups, instances when students articulated their expectations about the results and purpose of the experiment. Four of the six groups implied that they understood the purpose to be confirming the model.

One of these groups was analyzed in Ref. [35]. We selected two more of these four groups to evaluate how students' expectations interacted with their decision-making processes. These two groups were selected because the group members had similar expectations throughout the lab activity, which made descriptions of their expectations and corresponding decisions clear.

To analyze these videos, E. M. S. transcribed and annotated portions of the video where the students conveyed their expectations or made decisions based on their expectations. E. M. S. developed narratives that described students' expectations and consequent decisions and, while doing so, specifically focused on details surrounding students' decisions to engage in questionable research practices. From these narrative descriptions, we selected excerpts that conveyed students' expectations about the activity and how these expectations informed their engagement in questionable research practices. Following this process, M. M. S. and N. G. H. each read the selection of excerpts developed from the one of the narrative descriptions and deliberately watched the video to look for contradictory evidence to the presented descriptions. We made small changes to the descriptions surrounding excerpts (e.g., additional description of the situation) to accommodate possible alternative interpretations.

For the groups in the analyzed video, we then identified instances that the groups recorded their engagement in practices motivated by model confirming expectations in their electronic lab notebooks. Each time a student pressed a button to save revisions to the notebook, a copy of the entry was archived to allow for complete documentation of changes. We read through all documented revisions to the groups' lab notes to identify instances that they conveyed their model confirming expectations and any records of questionable research practices including those identified in Ref. [34] (see Table I).

IV. RESULTS

In this section, we present results that correspond to each of our research questions.

A. Students' expectations about labs

Students entering this course expressed expectations about the purpose of labs in similar ways as the students described in prior literature [13,25]. Table II shows the codes applied to students' written responses to "What do you believe is the purpose of a physics lab?" Most students entering the Physics and Engineering courses believed physics labs are about confirming, observing, or reinforcing the physics content that is introduced in other areas of the course, consistent with the idea that labs are an exercise in confirming models. Most students did not include experimentation skills in their responses.

The following examples highlight the characteristics of students' responses corresponding to confirmatory expectations about lab activities. The applied codes are provided in brackets following the coded statement. For example, one student wrote

TABLE I.   Coding scheme of questionable research practices recorded in groups' lab notes that was developed from groups' lab notes as described in Ref. [34]. Codes and descriptions are reproduced from Ref. [34] and examples are provided for clarity.

Category	Questionable research practice	Description	Example(s)
Subjective Interpretation	*Concerning results*	The distinguishability of the datasets is described as a concern or issue.	"*[The mismatch between data and the model was] a concern.*"
	Emotional response to data	Statements refer to students liking or disliking the results.	"*[We will change the experiment] if there is any unsatisfaction [with the results].*" "*We chose to go back to individual oscillations because we liked the low t' values of method 1 and we wanted to see if we could recreate that.*"
	Qualitative judgment of results	The distinguishability of the data or quality of the methods are judged qualitatively (e.g., good, bad, too small, too large, helpful, or an improvement) based on the results.	"*[We] liked the low t'.*" "*However, students t' increased, which is BAD, considering the theoretical expectation is that their periods are consistent.*"
Unjustified interpretation	*Claim of accuracy*	The accuracy of data, the method, or an instrument used to take data, are judged based on the test statistic value.	"*This was a very crude experiment which resulted in a high t' which means our experiment was not very accurate.*"
	Claim of systematic error	The distinguishability of the data sets is explained based on the presence or absence of systematic error, without describing the source.	"*If we could reduce our error... our t' should have been much lower*"
	Doubting statistics	The validity of statistical tools, like the test statistic or standard deviation/error is questioned without justification.	"*Our t' test is not representative of our data, this is because the standard error value is so large, giving us a lower t' value.*"
Purpose	*(Dis-)prove the model*	The purpose of the lab or intent of the group is explicitly or implicitly stated as to show that the model holds or breaks down.	"*Purpose: To determine the period of a pendulum is the same when released from 10 and 20 degrees.*" "*We will now try to be more precise by recording 10 consecutive periods, so that the periods are more alike.*"
Data manipulation	*Inflating uncertainty*	Statements demonstrate that students attempted to inflate their uncertainty, either through experimental decisions or manipulation of data.	"*Based on our t value, our values were possibly different [group found a t' of 1.67]. However, considering the unreasonably low uncertainty it is believed that our values were probably not different... Therefore we tried calculating the t value using 0.02 seconds as our uncertainty [group previously used uncertainty of 0.003 seconds]. After doing so, our t value equaled 0.8367 which would prove that our values were probably not different.*"

"*To show in the lab that the equations and concepts discussed in class [Supplement or demonstrate course material] indeed hold true and can be measured by you [Verify models or theories].*"

Another student recorded

"*To me, physics labs are meant as a supplement to the lecture and discussion sections [Supplement or demonstrate course materials]. They are meant to reinforce our knowledge of [Enhance understanding of concepts]*

and display some real-world applications [Real-world application of concepts] of the equations and concepts that we learn in class.*"

A large minority of students indicated that labs are used to develop experimentation skills, and often these students also responded with a confirmatory idea. For example, a student wrote

"*A physics lab is meant to provide a hands-on approach to what students are learning in a course [Supplement*

TABLE II. Codes applied to students' responses to "What do you believe is the purpose of a physics lab?" Italics indicate confirmation codes. Both introductory sequences include students who intend to pursue physics, engineering, and other majors; the labels indicate the most common major in the sequence. Students in group A answered the question prior to receiving instruction on the learning goals of the lab. Students in group B answered the question during or after receiving instruction on the learning goals of the lab. All students provided responses during the first lab session and prior to the lab activity described in this paper.

	Physics		Engineering	
	Group A	Group B	Group A	Group B
Code	($N = 93$)	($N = 39$)	($N = 342$)	($N = 35$)
Enhance understanding of concepts	52%	36%	44%	40%
Real-world application of concepts	23%	36%	37%	20%
Supplement or demonstrate course materials	44%	41%	61%	26%
Verify models or theories	19%	0%	8%	8%
Develop experimentation skills	31%	28%	13%	31%
Test models or theories	21%	23%	10%	14%
Problem solving or critical thinking	5%	0%	2%	3%
Team work	2%	3%	3%	0%
Only confirmation code(s)	43%	49%	72%	49%
At least one confirmation code[a]	82%	77%	88%	63%
No confirmation codes[a]	14%	18%	9%	26%

[a]Students whose responses were coded only as 'other' are not included, which accounts for totals less than 100%.

or demonstrate course materials]. Oftentimes, at least to students, physics can seem like a jumbled mess of equations with no real connection between any of them. Labs allows students a chance to see the practical applications [Real-world application of concepts] of what they are learning which will in turn provide them with a deeper understanding of the material [Enhance understanding of concepts]. Additionally, it is useful for students to acquire good laboratory and experimental techniques should they ever decide to go into research [Develop experimentation skills]."

These data suggest that students' expectations about the purpose of physics labs are commonly centered around the confirmation of models introduced in other areas of the course. The confirmation codes all require that models from lecture or textbooks hold in the lab environment. For example, real-world applications of physics concepts require that physical models sufficiently hold in applied contexts so that a student may learn about how the physical model applies.

As mentioned earlier, two instructors reported that they reordered the sequence of activities so that students wrote their responses to the prompt during or after an activity centered around the learning goals, potentially influencing their responses away from confirmation because the goals of the course focused on experimentation. Table II shows that Engineering Group B students' responses were less frequently coded with confirmation codes than Engineering Group A, however, there was no obvious difference between groups of physics students. The majority of students continued to include ideas about confirming phenomena in their descriptions, which suggests that

confirmatory expectations about physics labs are robust and not shifted with a brief instructional intervention.

B. Previous experiences with model confirmation labs

We suspected that the confirmatory expectations resulted from students' prior and concurrent experiences with physics and science labs and courses. In this section, we describe the broadly categorized experiences that students shared during the interviews, usually to contrast their lab experiences in physics labs centered around experimentation skills.

All interviewed students ($N = 10$) reflected on components of their prior or concurrent lab experiences in their high school or nonphysics college courses. The students' descriptions about their lab experiences indicated variation in the amount of structure and available control over the experimental design. However, most students' descriptions indicated that other labs incorporate aspects of conceptual reinforcement, most often in the form of confirming models.

The most frequent ($n = 9$) comment that students shared about their previous experiences was that the aim of their lab experiences was to obtain a specific result. For example, one student described "*like in high school you knew that the lab was entirely designed to get you to come to a specific conclusion or specific result. What you'd do is just figure out what that was and then write it at the beginning. That was our goal for the whole time, and then you look like [snaps fingers] then you look like you've understood the whole content from the whole perspective and you'd get your A and you'd leave class.*"

Most students ($n = 9$) did not speak about the instructional goals of their nonphysics college labs, however, one student described their concurrent college-level labs as having different objectives from the physics labs. The student provided an example: "*That makes it a lot different from like any other lab I've been in… when I'm in chem lab, it's not like that at all. It's hard core and it's like you have to get the results that you want. And if you don't get the results you want, it actually kind of messes you up, which is why I actually hate, uh, I hated it. Because if you're finding something different from what you're supposed to be getting, you should be able to present that.*" This student's reflections suggest that college labs may also contribute to students experiencing lab courses as activities in confirmation and, possibly, provide negative experiences in the process of obtaining confirmatory results.

Most students ($n = 7$) compared the confirmation aspects of other lab experiences to the non-confirmation goals of the physics labs. For example, a student provided a succinct comparison that "*we haven't, for example, been verifying that phenomena we talk about in class exist, which is what a lot of other labs have done, have been.*" Another student contrasted his current experience in the physics labs with prior lab experiences, however, also expressed a preference for lab activities that were not simply centered around confirmation. He stated: "*I guess I really like the sort of, like, different direction than… most labs. Like at this point I've, like, done a lab to show that, like, conservation of momentum works, like, experimentally, like, for two years in a row. So I'm glad it wasn't another semester of that because, like, I always thought those labs were a waste of time. But I think, like, taking the more, like, step-back approach has been, like, kind of nice. [Inaudible] Comparing models, proving slash disproving that this model holds true is kind of nicer.*"

The amount of structure that students experience in other labs may vary greatly among lab experiences that include a specific result as an objective. Two students described experiences in labs with relatively little structure, however, most students did not provide enough information to gauge the amount of structure within their prior lab experiences. One student described their high school physics labs to be where "*as we would walk in the room, the teacher would write the question on the board like, you know, prove the gravitational constant g is close to 9.8 m/s^2 or something like that. And she wouldn't give us anything else, she'd put a pile of materials on the table and be like 'here's some things you can use for your lab.' And we would be expected to come up with the objective, the procedure, hypothesis, the whole scientific method.*" In this lab the students were responsible for the experimental design and analysis decisions, however, the student recalled the purpose of the lab activity to be, nonetheless, centered around confirmation.

The other student with an unstructured lab experience described prior lab experiences without identifying confirmation goals within those experiences. The student recalled experiences with little structure and without objectives provided where "*[he] had a teacher in high school that would literally just tell us 'design a lab about energy' and then no one knew anything about what they were doing, and it was just a disaster.*" Without an explicit objective for the lab and due to the lack of structure, the student perceived the experience to be unproductive. This appeared to be an uncommon experience among the interviewed students.

The only student who did not explicitly mention confirmation as an aspect of any prior lab experiences also did not describe non-confirmation experiences. Instead, he focused his descriptions on the differences in equipment among his lab experiences. Throughout these descriptions, he implied that labs are a means by which to observe phenomena; many of his statements suggested that this was a common lab experience for him. At the time of the interview, he was enrolled in an electricity and magnetism physics lab with instructional goals centered around reinforcing content introduced in the lecture portion of the course.

Students' descriptions of their experiences suggest that their other experiences in instructional labs may align with, and probably influence, their incoming expectations. Labs that are designed to confirm a model through experimentation appear to be common among the interviewed students, however, the amount of structure in these experiences may greatly vary from highly guided to open ended. Most students contrasted their experiences in the physics labs with their experiences in other instructional labs, which suggests that students may regularly identify differences in instructional goals between different lab environments.

C. Model confirming expectations influence engagement in questionable research practices

In this section, we present evidence that suggests students' expectations that labs are activities in model confirmation lead them to engage in questionable research practices. We present examples of how students' expectations about the purpose of labs manifest in their first activity. We then provide examples of groups that exhibited confirmatory expectations. These examples are not exhaustive but demonstrate some of the ways in which confirmatory expectations manifest in lab activities, especially when instructional goals are in tension with students' expectations for the activity. We focus on two groups that convey their expectations about the purpose of the lab activity to be an exercise in model confirmation. An illustrative description of an additional group's decision making during the activity is provided in Ref. [35].

1. Conveying expectations about the purpose of lab

We describe two groups: Bria, Max, and Phoebe and Ivy and Charlie [41]. Both groups engaged in conversations and made decisions that revealed their expectations during the lab activity. Both groups regularly conveyed their intent

to prove that the model holds in the lab, a questionable research practice that was also identified through our analysis of groups' lab notes [34]. However, unlike the group described in Ref. [35], neither group began the lab with an explicit discussion of the purpose of the activity.

The groups were in different lab sections and had different instructors. Prior to students' investigations, the instructors addressed the whole class with explicit messaging about their expectations for students' engagement with the activity that hinted that grading was not tied to obtaining a particular result.

Bria, Max, and Phoebe's instructor explicitly discussed the goal of the lab course. At the beginning of the lab session, he led a discussion that began and ended with comments about the purpose of the activity. After calling the attention of the students, he outlined the goal of the lab activity: "*you will be attempting to determine whether there is a difference, that you can measure, of [a] pendulum started at 10 degrees and a pendulum started—the period of a pendulum started at 10 degrees or the period of a pendulum started at 20 degrees.*"

At the end of the whole class discussion, Bria, Max, and Phoebe's instructor, again, summarized the purpose of the lab: "*you all need to convince me that either there's a difference that you've measured between 10 and 20 degrees of the pendulum or there is no such difference that you can measure. I want you to sell me the package. Is there a difference between 10 and 20 degrees of the period of the pendulum? Alright? And you can do that by whatever means you see necessary.*" The instructor's messaging conveyed the possibility of multiple outcomes to the students and that the evidence from the process was more important than the outcome.

Ivy and Charlie's instructor did not explicitly discuss multiple possible outcomes for the experiment, however, emphasized process over results throughout his introduction to the activity. The instructor stated that groups' lab notes should include descriptions of what they did, why they did it, and follow-up investigations. Additionally, the instructor was careful in language choices around the model and used phrasing such as "*comparing measurements of amplitudes at 10 and 20 degrees*" and "*what we're testing is this equation [pointing to $T = 2\pi\sqrt{L/g}$ written on the front whiteboard].*"

Bria, Max, and Phoebe implicitly conveyed their model confirming expectations through discussions about expected results and selective analysis of data. They began the lab by developing a procedure for their data collection and, at length, discussed the formatting and content of their lab notes. After nearly fifteen minutes, the students began to collect data and never mentioned expectations about results prior to collecting data.

After the students followed their procedure for collecting data at 10 degrees, they transitioned to collecting data for the period at 20 degrees. When Phoebe obtained a measurement that may have included a mistaken count, Bria compared this single data point at 20 degrees to the dataset at 10 degrees and asked her group "*they're supposed to be the same, they're not supposed to be different, right?*" Through this statement, Bria first exhibited confirmatory expectations: She expected measurements of the period at 10 and 20 degrees to be the same. Neither Max nor Phoebe directly challenged Bria's assertion, though Max remarked "*theoretically, right?,*" which implicitly suggests that they shared Bria's expectations about results. Unlike the students in Ref. [35], they did not explicitly state that their perception about the purpose of the lab is to demonstrate that the periods are the same.

Ivy and Charlie are similar to Bria, Max, and Phoebe in that they never explicitly discussed their expectations but implicitly suggested their expectations through interpretations of their data. However, they were primarily driven by a desire to finish the lab and therefore, prioritized plans that expedited data collection. Their first iteration involved explicitly planning measurements that Charlie explained would "*save us a little bit of time because I don't want to be here forever.*" However, neither student discussed expectations about results until nearly halfway through the lab period. After obtaining data at both 10 and 20 degrees, Charlie subjectively interpreted their data by comparing the means and commented that "*yes, I think this is good, oh my god, they're the same even with like the hundreds*" to which Ivy replied "*that's good.*" Without performing any statistical comparisons, Charlie concluded "*that's like very significant, what the, I guess, it's a good thing; that's kind of insane!*" These subjective interpretations of the means of their datasets suggest that both students expected the periods to be the same at different angular displacements. In tandem with conveying their expectations, they also engaged in a questionable research practice surrounding subjective interpretation of data.

2. Model confirming expectations lead to questionable research practices

Here, we provide examples from the video when the groups' confirmatory expectations led them to engage in other questionable research practices. No groups engaged in practices that are flagrant examples of research misconduct and many of the students' decisions were reasonable given their model confirming expectations. These examples highlight the conflict of students' confirmatory expectations with the instructional goals of the lab, leading students to engage in questionable research practices.

Bria, Max, and Phoebe.—After several trials, Bria, Max, and Phoebe obtained around two for their t' value about which Phoebe remarked "*is like smack dab in the middle of who knows what's going on.*" Bria, Max, and Phoebe followed this by combining their data from the different measuring methods (measurements from the highest and lowest points of the swing), as an explicit attempt to obtain

a lower value of t'. However, this method did not result in t' less than one, so they began to selectively analyze their data. First, the group rechecked a subset of their data. They chose to analyze data from measurements from the lowest point of the swing *because* these measurements had greater precision and avoided what they referred to as a 'bad' data point. The contrast between the responsible conduct of research practice to use measurements with more precision and the questionable research practice to obtain better agreement suggests students engage in behaviors that align with instructional intent in addition to holding confirmatory expectations. In their perspectives, their selective analysis of their data turned out to be unhelpful because t' remained in the inconclusive range. Jokingly, Phoebe summarized "*that didn't work so we throw that data out and start over.*"

As the students prepared for acquiring more data, they commented that "*[they]'re the ones measuring, so we're responsible for the failure of this experiment,*" which communicated their view of a failed experiment as one in which their t' remained above one. They sustained their confirmatory expectations. As they obtained additional data, the students found that the standard errors of the means of their new data sets were equal, and expressed their excitement with exclamations of "*that way they're the same!,*" "*solid!,*" and "*I hope this makes a difference!.*" However, their celebrations were brief as they completed their calculation to find that with additional data, the value of t' had, again, increased. They responded with "*no!,*" "*but why did that happen?,*" and "*that shouldn't have happened,*" communicating their frustration with results because they believed their data were improved over the previous iterations. These responses are examples of the group's emotional responses to results, perhaps a questionable research practice when emotions are unchecked, and motivated their decisions to further engage in selective data comparisons.

Bria provided an opportunity for the group to shift their experimentation goals by asking Phoebe and Max "*but seriously, why is that number so big?*" Rather than reevaluating their approach to the lab, the students decided to recalculate t' by rounding decimal places in various ways but returned to including the entire dataset; each student performed calculations with different rounding choices, reducing the precision obtained from repeated trials. After doing so, one student announced that she "*got a better one but not below one*" and followed up with a plan that "*if these numbers don't come out, then let's just do the lows.*" The 'lows' were the dataset from measuring the period of the pendulum at the lowest point of the swing and did not include the 'bad' data point. They continued to analyze various rounding choices in calculations of t'. One student commented that "*when I took off a bunch of decimal places, I still got the 1.99.*" The students became increasingly vocal in their frustration with the unexpectedly high values of t' and asked "*what are we doing wrong?*"

Upon reflecting on the group's result, Bria expressed some openness to alternative ideas: "*they're definitely supposed to be the same period, right? I'm not going crazy*" to which Phoebe replied "*I sure thought so!*" This exchange highlights how Bria and Phoebe's shared model confirming expectations were dominant throughout most of the lab activity. Bria continued on to claim that "*the data we take shows [that they're the same]*" and Phoebe chimed in support by discussing that the means "*are super close*" for the periods at 10 and 20 degrees. Phoebe concluded the conversation about their expectations by explicitly pointing out their results did not contradict their expectations: "*And it's not like our t value is saying that they are different. It's just not saying anything.*"

Shortly after and with only 20 min left in the lab period, Bria, Max, and Phoebe heard two other groups yell out that they obtained t' values of 3.7 and 3.8. Hearing this, Phoebe turned to Bria and remarked that their most recent value of t' of "*2.9 is really close to 3!*" Bria responded "*I just want it to go one way or the other; at this point I just don't care.*" When the students heard that other groups obtained values indicating distinguishable periods, they immediately opened to pursuing a new goal to finish the lab activity. They desired a conclusion that indicated the periods either were or were not distinguishable rather than inconclusive results.

However, simultaneous to the shifted goal of disconfirming the model through experimental evidence, Bria maintained her goal of confirming a model. She provided an explanation that the periods are mathematically identical, but measurements were limited by other forces. Bria suggested to Max and Phoebe that drag forces cause the periods to be measurably different and that the model ($T = 2\pi\sqrt{L/g}$) does not account for the drag forces and explains "*that's why they made us do like 10 and 50*" in high school.

Bria's explanation incorporates, using the language of our codes developed from groups' lab notes, an example of an unjustified interpretation by making a claim of systematic error. Max and Phoebe did not propose alternative hypotheses, which suggests that they agreed with Bria's explanation.

Ivy and Charlie.—Throughout the lab activity, Ivy and Charlie were frequently off task. The majority of their time was not used to engage in the lab activity, and prodding from authoritative sources (i.e., lab instructions and instructor) was required before they made decisions in the lab activity. Ivy and Charlie's discussions strongly suggest that they expected and intended to leave after obtaining confirmatory results.

After the group obtained their first value for t', Charlie shared with Ivy that "*it is indistinguishable because it's 0.65.*" Ivy replied by stating the conclusion: "*so we got the right number we need.*" This exchange highlights the persistence of the students' expectations that they were

confirming the model that was provided to them by the lab instructions. After obtaining the t' value that confirmed their expectations, Ivy made a bold statement that "*well maybe at the end of this, we'll be like 'oh our t' was under one so we're calling it a day and we're leaving'.*" The students drew a conclusion for the entire lab from a single iteration that they confirmed the model.

Charlie and Ivy's model confirming expectations continued throughout a whole class discussion. The instructor prompted groups to raise their hands to indicate the interpretation of their latest calculation of t'. After viewing many groups raising their hands for t' in the distinguishable and inconclusive ranges, Charlie whispered to Ivy "*how are we so lucky?*" In light of contradictory evidence from peers in other groups, the students viewed their result as correct and did not consider other experimental outcomes.

After the whole class discussion, Charlie shared with Ivy an explanation for systematic differences in the period of the pendulum that other groups had uncovered. She commented that "*the trick is when you release it, you don't push on the angle*" and explained that "*when I release it, I always just do that [carefully releases pendulum without a push], like a really quick release. Some people, like, do that [gives a push to the pendulum while releasing it].*" Ivy appeared to buy Charlie's explanation by commenting "*now I'm depressed.*" Charlie engaged in the questionable research practice of making an unsubstantiated claim of other groups' systematic errors that resulted from inferior measuring techniques of their peers without evaluating how other groups were obtaining their measurements.

Eventually, the students followed the instructor's suggestions to try a much larger angle. They decided to compare the period at 60 degrees to their previous datasets at 10 and 20 degrees, which the instructor stated would yield different periods. Without performing any statistical analyses of the data, Charlie explained to the instructor that "*I guess I mean part of the reason why is because we, technically we really have air resistance so that can be a confounding variable. When you're using the formula, you're only accounting for, because air resistance is an opposing force that causes it to deccelerate so that, I guess the bigger the angle, the more air resistance.*" Charlie recognized that the instructor expected the period to be different at 60 degrees, and so, reacted by explaining why the periods will be different prior to performing any analyses. Charlie's explanation is unsubstantiated and, is again, an unjustified claim of a systematic effect.

However as the instructor began to walk away, Charlie returned to her original description of the systematic error involving a push to the pendulum bob at release. She explained to the instructor that the difference with the measurements of the period at 60 degrees was that "*I guess maybe I accidentally pushed it a little bit too because it's bigger and it's harder to control, so I probably pushed it a little bit, maybe, I don't know. If we had more data points, then I'd guess it would be more accurate.*"

With a few minutes left in the lab period, the instructor called the whole class together for a discussion. During the discussion, Ivy was paying attention and Charlie performed calculations of t' to compare 10 and 60 degrees. When Charlie obtained a t' value, Ivy had just heard that the model contained a small angle approximation. Charlie commented that "*oh [t']'s a big one, that's weird*" to which Ivy replied "*wow, is that our t'?*" Charlie explained that "*yes, it's so weird*" and Ivy provided the explanation that she had heard "*it's because it's not a small angle.*" However, Charlie did not recognize what Ivy has conveyed and replied "*but it's just so strange.*"

3. Lab note records of decisions

All students were instructed to record and justify their decisions in their electronic lab notes. These notes can be cross referenced with the video analysis to understand what the students decided (not) to communicate to their instructors. Furthermore, the software periodically tracks revisions and updates to entries, providing more detail than the analysis of submitted lab notes as in Ref. [34].

Bria, Max, and Phoebe.—Bria, Max, and Phoebe's submitted lab notes contained no information about their initial intent to confirm the model provided to them. Furthermore, there are no statements in their lab notes that were coded as questionable research practices, and their recorded conclusion was the periods were distinguishable at 10 and 20 degrees. Their lab notes were not included in the fraction of groups that exhibited at least one questionable research practice in Ref. [34]. Therefore, we believe it is reasonable to assume that the frequencies reported in Ref. [34] are lower estimates.

However, the submitted version of their notes was revised from an earlier saved version. In the earlier version of the notes, they stated that the purpose of the lab was to "*determine the period of a pendulum is the same when released from 10 and 20 degrees.*" By the time that they submitted their lab notes, they, perhaps, recognized that this was not the objective of the lab and changed their language to reflect what they perceived to be the real intent.

Charlie and Ivy.—Charlie and Ivy recorded several statements in their lab notes that conveyed their model confirming expectations, however, there was little record of their engagement in the questionable research practices seen in the video. Early in their lab notes, they recorded that "*if the uncertainty is still high and our t' comes out over 1, we will use 10 periods instead of 5 and try again.*" In this plan, they expressed that they expected t' to be less than one and that following up with improved measurements is only necessary if there was not agreement. However, later parts of the lab notes maintained objectivity in language choices and did not convey their sustained intent

to confirm the model nor did they record their explanations of systematic effects.

Similar to Bria, Max, and Phoebe, Charlie and Ivy modified their lab notes as they completed the lab. One of their modifications was removing an interpretation for their second t' value. Originally they interpreted a t' of 0.3552 as a low value and that "*there is enough evidence to suggest that angle doesn't affect the period of a pendulum.*" After they found contradictory results, they removed this statement. Interpretations of results change in light of new, contradictory evidence, however, the presence and subsequent removal of this statement suggests that the students did not want to disclose their initial understanding of the task to their instructor.

V. DISCUSSION

In this exploratory study, we aimed to answer three research questions about (i) students' expectations about the purpose of physics labs, (ii) how those expectations relate to their prior lab experiences, and (iii) how those expectations relate to students' engagement in questionable research practices in instructional physics labs.

Related to our first research question, our results suggest that many students entered their lab course expecting that models introduced in lecture would be confirmed experimentally during the lab. These expectations were found to be common in two distinct populations of students in introductory physics courses (namely, students intending to pursue physics and engineering majors) and appeared resilient to explicit instruction about non-confirmatory goals and objectives. Students' beliefs about the nature of science have been similarly found to be resilient to simple instructional interventions [42]. However, students' expectations during instructional activities are typically considered to be flexible and responsive to context [20,43]. Future work should further probe the resiliency and flexibility of students' model confirming expectations in lab.

Related to our second research question, we found that students' confirmatory expectations came from many previous and current lab experiences with confirmatory aims. In our analysis, we did not explore other possible sources. For example, students' confirmatory expectations are similar to the portrayal of science in popular culture. Many recent news articles portray experiment as a means to verify theory. For example, announcements of LIGO's initial gravitational waves detection and the recent imaging of a black hole were both advertised as confirmation of Einstein's theories (e.g., Refs. [44–51]). The experimental feats were secondary to the fact that Einstein had been proven right. Such messaging may implicitly (or in some cases explicitly) suggest that confirmation is not only regular practice but the primary goal of experimental science, influencing students' understanding of what it means to engage in experiment. Future work, therefore, should seek to resolve the degree to which students' beliefs

about instructional physics labs described here and elsewhere [11–13] stem from sources beyond their previous labs and should seek to compare these understandings to students' broader beliefs about experimental physics.

Related to our third research question, we found through video analysis of groups conveying confirmatory expectations that students' confirmatory expectations manifest with students' engagement in questionable research practices. That is, students engaged in questionable research practices (such as selectively analyzing data, ignoring contradictory results, manipulating data to change the value of a statistic, responding emotionally to unexpected results, and forming unjustified explanations about the situation) in attempts to confirm canonical or given models. While some students conveyed their model confirming expectations through their lab notes, most model confirming decisions were unrecorded. This result suggests that previous work underreported students' engagement in questionable research practices [34] and that engagement in questionable research practices may be more common than observed in the products submitted to an instructor. However, we did not analyze video of groups that did not have confirmatory expectations. Future work should aim to develop contradictory narratives to our presented results to test our interpretations.

Our data do not allow us to claim that confirmatory expectations necessarily lead students to engage in questionable research practices. For example, not confirming an expected result may suggest an error was made, and productively send the student into a troubleshooting mode. Nor do we claim that these behaviors were necessarily inappropriate or unethical. Indeed, many expert experimentalists may engage in behaviors similar to those we identified as questionable research practices, while conforming to rigorous and ethical research practices. For example, emotional reactions to results of experiments are a regular part of research [52,53]. However, expert scientists are cautiously skeptical of results that agree with their expectations and are motivated to understand why other results violate their expectations [17]. Our results instead illuminate the tension between students' expectations about labs and the instructional goals that aim to engage students in the practices and processes of experimental physics. Our research suggests that students are developing appropriate methods for reacting to expected and unexpected experimental results and that many instructional labs may provide students with experiences that shut down informed skepticism to promote model confirmation.

We also are not claiming that lab activities with a known or expected outcome should never be conducted. The data suggest that care must be taken for how lab activities are oriented, framed, and contextualized. For example, Allie and colleagues described an experiment that aimed to verify a value for the acceleration due to gravity with a simple pendulum [54]. The activity, however, was contextualized by

a hypothetical group of researchers who measured a value that conflicted with the canonical value. The activity, therefore, oriented students such that they were expected to critically evaluate the experimental methods and equipment to resolve the conflict. As another example, Marasco described an experiment aiming to test whether the energy of a bouncing ball is primarily transferred during the bounce or due to drag during the fall [55]. Most students expected the energy to be transferred during the bounce, but subtle choices in the data analysis process may produce results consistent with energy transfer due to drag or even energy gains to the system during the fall. Conflicts between students' expectations and their results set up productive discussions about the data and analysis methods. These examples demonstrate that lab activities that appear confirmatory, but actually yield surprising or conflicting results, may be productive for students' engagement and learning [56–58]. Our data, however, support the claim that lab activities developed with the intent for a student to obtain a particular outcome, without careful attention to evaluating a scientific process or model, may lead to problematic biases, behaviors, and beliefs.

The data also suggest that confirmatory expectations may motivate other behaviors and ways of framing the activities in introductory labs. We explore two of these: relationship with a 'desire to be done' and responses to conflicting evidence.

A. Desire to be done may often align with confirmatory expectations

The examples of students' decisions and discussions during lab highlight that model confirmation is not the only motivation for engaging in questionable research practices. Students' desire to be done with their experiment also manifests in a variety of questionable research practices. In a confirmatory lab, the activity is complete after experimental results confirm a model. A common feature of both groups was a belief that they would be done with the activity after obtaining a conclusive, confirmatory result.

This alignment was particularly prominent throughout Ivy and Charlie's exchanges during the lab. Their desire to finish the lab is obvious from the beginning of their investigation and dominates their desire to confirm results. In their perspectives, the signal that they finished the investigation is that they obtained confirmatory results. They viewed their interactions with the lab instructor to be exasperating because the instructor kept pushing them to obtain improved precision, but they believed that they had completed the activity.

B. Conflicting evidence may lead to new goals or to reinforce confirmation

Between the two groups analyzed on video, there is a similar event when the students learned that other groups obtained results that showed the periods were distinguishable. Bria, Max, and Phoebe heard students yell out their results across the room, and Charlie and Ivy saw many groups raise their hands during a whole class discussion. The groups reacted differently to the information.

Bria, Max, and Phoebe reacted by pointing out that their results were nearly distinguishable and rapidly shifted their goal to obtaining any conclusive result. Their model confirming expectations were not relevant when they learned of other groups' results that pointed to distinguishable periods. Obtaining a conclusive result became a higher priority than confirming the model. However, when Charlie and Ivy confronted similar information, they viewed themselves as "*lucky*" because they obtained confirmatory results, unlike several of their classmates. Their expectations did not shift as a result of new information about others' results and continued to view their results as successful.

There are two noticeable differences between these situations: (i) the result that the groups had when they received this information and (ii) the way in which they received the information. Bria, Max, and Phoebe had been consistently obtaining inconclusive results; their frustration with these results may have opened the space for alternative conclusions. In contrast, Ivy and Charlie obtained results that indicated the periods were indistinguishable, confirming their prior expectations. This experimental result may have influenced them to believe that they carried out the experiment better than their peers, reinforcing their model confirming expectations.

Furthermore, Bria, Max, and Phoebe received the conflicting evidence from peers loudly communicating across the lab room and without input from the instructor. But Charlie and Ivy received their information through a whole class discussion where people raised their hands for the conclusion that their group was drawing from a t' value. A similar instructor-led discussion took place in Bria, Max, and Phoebe's lab and, similarly, did not cause any change in their confirmatory goals. Their behavior changed when another student, unsolicited and with conviction, announced that the periods were distinguishable. The ability for students to convey conviction in their results that conflict with prior expectations may influence how groups take in and interpret new, contradictory evidence. Alternatively, the class discussion, where multiple groups report conclusions that span distinguishable and indistinguishable measurements of the periods at 10 and 20 degrees, may maintain ambiguity of the results that students may expect.

C. Implications for instruction

The results in this study shed more light on the issues with traditional labs. While traditional labs are typically criticized for the extensive amount of structure, our data suggest that the confirmatory nature may be just as or more

problematic. Our results contribute to the debate about whether the goals of introductory labs should focus on conceptual understanding, developing experimentation skills, or both. The results here, supported by data on students' attitudes towards experimental physics [59], imply that lab activities also cannot do both.

Because of the prevalence of model confirming expectations and a potential link to students' broader beliefs about the nature of science, an activity such as Pendulum for Pros may serve as an opportunity to explicitly discuss the nature of experimentation. For example, following the Pendulum for Pros activity, instructors could ask students to reflect on their experiences and how their actions were similar to or different from those of expert experimental scientists.

We also believe this activity serves as an opportunity to teach responsible conduct of research (RCR) by providing an opportunity for students to discuss how to mitigate inevitable expectations and implicit biases. Developing effective training in ethics and RCR is a key goal of most (if not all) scientific agencies and associations [60–63] but effective training programs remain rare [64,65]. A meta-analysis of studies evaluating RCR instruction found that instruction is, at best, moderately effective and the most effective instruction included opportunities for students to apply decision-making skills to authentic cases [66].

Students' tendency to engage in questionable research practices due to their model confirming expectations may create a time-for-telling [67] opportunity to teach RCR. Many students engaged in behaviors well aligned with rigorous and responsible research practices. For example, Bria, Max, and Phoebe developed and implemented methods to reduce the uncertainty of their measurements of the period despite their aim to confirm. These behaviors suggest that students are well positioned to reflect on their decisions and motivations for those decisions. We have developed an intervention at the beginning of the next lab activity for students to reflect on and discuss their decisions in the context of their expectations and biases and connect to larger issues of RCR and the refinement of scientific models through experimentation. In future work, we plan to analyze whether this intervention is effective in teaching students about RCR and, possibly, shifting their expectations about and behaviors during lab activities.

Alternatively, this activity may have simply relocated how students focus on confirmation. At the end of one of the episodes, Bria re-aligned to aim to confirm a model with the angle dependence and demonstrate that the periods at the two angles are distinguishable. An alternate intervention for teaching RCR may be to use investigations that allow for multiple possible approaches and outcomes. For example, students can evaluate the degree to which 'stretchy' objects from home obey Hooke's law. An electricity and magnetism lab could evaluate the physical properties of LEDs [68] well before students receive relevant instruction.

VI. CONCLUSIONS

As instructors and curriculum developers look to improve lab instruction, attention needs to be given to students' expectations about labs and the implicit messages they may be receiving. At the institution in this study, the instructional messaging about the purpose of the lab curriculum was explicit prior to the lab activity. However, students' expectations were resilient to strong instructional messaging and caused them to engage in questionable research practices. The explicit instructional aim to verify, prove, or confirm a physical model in physics labs may problematically communicate to students that science is an exercise in confirmation.

APPENDIX A: INCOMING EXPECTATIONS CODING SCHEME

The following codes were developed from students' responses to "What do you believe is the purpose of a physics lab?" Several codes are similar to and informed by those from Ref. [12].

Enhance understanding of concepts: Labs help develop an understanding of physics (e.g., concepts, equations, laws, etc.) and/or help develop physical intuition with physics.

Real-world application of concepts: Labs demonstrate how physics (e.g., concepts, equations, laws, etc.) applies to real-life situations and/or show a practical use for physics. Students apply or use the principles in the real world. It is not, simply, a demonstration of course materials but requires an application or use of the physics in the "real world."

Supplement or demonstrate course materials: Labs are a supplement to lecture, textbook, homework, and/or discussion parts of the course. Demonstrations of the materials from class are included.

Verify models or theories: Labs are exercises in demonstrating, verifying, and/or confirming that a model or theory holds, works, and/or is true.

Develop experimentation skills: Labs help develop skills in experimentation (e.g., data analysis, how to develop procedures, etc.). This code is broad and applied liberally.

Test models or theories: Labs are exercises in testing theories (without explicit tone of confirmation).

Problem solving or critical thinking: Labs help develop problem solving, critical thinking, and/or thinking like a scientist skills.

Team work: Labs help develop cooperation or collaboration skills.

Other: Responses or portions of responses that do not fit into the preceding codes.

APPENDIX B: ABBREVIATED INTERVIEW PROTOCOL

Specific instructions for interviewers have been removed from this version of the interview protocol. Many of these instructions included suggested follow-up questions, specific information to avoid telling interviewees, and when the interviewer could ask questions of interest that were brought up by the students. Italicized questions indicate the questions that were consistent among interviewers. When there was additional time allocated for each section of the interview, then interviewers asked the follow-up questions that are not in italics.

Icebreakers
   (a) *What do you think you're supposed to be learning in your labs?*
   (b) *What do you think you're actually learning in your labs?*

Section I: Recall using lab notes
   (a) *Your group's lab notes from Pendulum for Pros are printed out for you. Use your lab notes to describe your group's decisions and processes throughout Pendulum for Pros. Please read aloud sections of the notes and then describe what your group was doing, what discussions your group had about those decisions, and why you settled on those decisions.*
   (b) *What do you think you learned from the Pendulum lab?*
      (i) *How might you use this in future classes, research, work, etc.?*
      (ii) *Have you used these lessons in any of the subsequent labs?*

Section II: Recall using lab handout
   (a) *The handout that you received during Pendulum for Pros is printed out for you. Let's focus on the activities that involve comparing the periods at 10 and 20 degrees, which are marked on the handout. As you read through the activity, discuss what you believe are the intentions of the activity and why you were asked to complete those activities. If you're able to recall, also discuss how you interpreted the activities during the lab session.*

   Prompts for while students are answering the question:
      (i) Do you recall what you believe the purpose of [the activity] was while you completed the lab?
      (ii) Now, what do you believe [the activity] is asking students to do?
   (b) *Do you recall your group discussing the purpose of comparing the pendulum at 10 and 20 degrees? If so, what did you discuss? When did you discuss it?*
      (i) *Did your group's idea about the purpose shift or change throughout the activity?*
      (ii) If <u>no mention</u> of iterating on measurements: *A focus of this lab was on iterating and improving measurements. Why do you think that was a focus of the lab?*

Section III: Follow-up questions
   (a) *We've found that several groups set out to prove that the periods of the pendulum are the same at 10 and 20 degrees. Why do you think several groups set out to demonstrate that the periods are the same?*
      (i) What sorts of situations might be good to set out trying to demonstrate that something holds up? What situations might this be a bad thing to do? Why?
   (b) *Is there anything you'd like to share with me about the labs this semester (or last semester)?*
   (c) *What questions do you have for me about what we've discussed, what my role is, or anything else?*

---

[1] V. K. Otero and D. E. Meltzer, 100 Years of attempts to transform physics education, Phys. Teach. **54,** 523 (2016).

[2] V. K. Otero and D. E. Meltzer, The past and future of physics education reform, Phys. Today **70,** No. 5, 50 (2017).

[3] E. Etkina, A. Van Heuvelen, S. White-Brahmia, D. T. Brookes, M. Gentile, S. Murthy, D. Rosengrant, and A. Warren, Scientific abilities and their assessment, Phys. Rev. ST Phys. Educ. Res. **2,** 020103 (2006).

[4] R. Lippmann Kung, Teaching the concepts of measurement: An example of a concept-based laboratory course, Am. J. Phys. **73,** 771 (2005).

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

[6] AAPT Committee on Laboratories, *AAPT Recommendations for the Undergraduate Physics Laboratory Curriculum*, (American Association of Physics Teachers, College Park, MD, 2015).

[7] National Research Council, *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*, edited by S. R. Singer, N. R. Nielsen, and H. A. Schweingruber (National Academies Press, Washington, DC, 2012).

[8] Joint Task Force on Undergraduate Physics Programs, *Phys21: Preparing Physics Students for 21st Century Careers*, (American Physical Society and American Association of Physics Teachers, College Park, MD, 2016).

[9] Next Generation Science Standards Lead States, *Next Generation Science Standard: For States, By States* (National Academies Press, Washington, DC, 2013).

[10] 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* (United States Office of Science and Technology Policy, Washington, DC, 2012).

[11] D. Hu, B. M. Zwickl, B. R. Wilcox, and H. J. Lewandowski, Qualitative investigation of students' views about experimental physics, Phys. Rev. Phys. Educ. Res. 13, 020134 (2017).

[12] D. Hu and B. Zwickl, Examining students' personal epistemology: The role of physics experiments and relation with theory, in *Proccedings of the 2017 Physics Education Research Conference, Cincinnati, OH*, edited by L. Ding, A. L. Traxler, and Y. Cao (AIP, New York, 2017), pp. 11–14.

[13] R. Lippmann, students' understanding of measurement and uncertainty in the physics laboratory: Social construction, underlying concepts, and quantitative analysis, Ph.D. Thesis, University of Maryland, 2003.

[14] P. S. Oh and S. J. Oh, What teachers of science need to know about models: An overview, Int. J. Sci. Educ. 33, 1109 (2011).

[15] S. Manthey and E. Brewe, Toward University Modeling InstructionBiology: Adapting curricular frameworks from physics to biology, CBE Life Sci. Educ. 12, 206 (2013).

[16] B. M. Zwickl, D. Hu, N. Finkelstein, and H. J. Lewandowski, Model-based reasoning in the physics laboratory: Framework and initial results, Phys. Rev. ST Phys. Educ. Res. 11, 020113 (2015).

[17] C. Wieman, Comparative cognitive task analyses of experimental science and instructional laboratory courses, Phys. Teach. 53, 349 (2015).

[18] D. R. Dounas-Frazer, L. Ríos, B. Pollard, J. T. Stanley, and H. J. Lewandowski, Characterizing lab instructors' self-reported learning goals to inform development of an experimental modeling skills assessment, Phys. Rev. Phys. Educ. Res. 14, 020118 (2018).

[19] E. F. Redish, J. M. Saul, and R. N. Steinberg, Student expectations in introductory physics, Am. J. Phys. 66, 212 (1998).

[20] R. E. Scherr and D. Hammer, Student behavior and epistemological framing: Examples from collaborative active-learning activities in physics, Cognit. Instr. 27, 147 (2009).

[21] T. J. Bing and E. F. Redish, Analyzing problem solving using math in physics: Epistemological framing via warrants, Phys. Rev. ST Phys. Educ. Res. 5, 020108 (2009).

[22] L. Lising and A. Elby, The impact of epistemology on learning: A case study from introductory physics, Am. J. Phys. 73, 372 (2005).

[23] D. Hu and N. S. Rebello, Shifting college students' epistemological framing using hypothetical debate problems, Phys. Res. ST Phys. Educ. Res. 10, 010117 (2014).

[24] P. W. Irving, M. S. Martinuk, and E. C. Sayre, Transitions in students' epistemic framing along two axes, Phys. Res. ST Phys. Educ. Res. 9, 010111 (2013).

[25] D. Hu and B. Zwickl, Examining students' views about validity of experiments: From introductory to Ph.D. students, Phys. Rev. Phys. Educ. Res. 14, 010121 (2018).

[26] M. Jeng, A selected history of expectation bias in physics, Am. J. Phys. 74, 578 (2006).

[27] Federal Policy on Research Misconduct Federal Register: December 6, 2000 (Volume 65, Number 235), pp. 76260–76264.

[28] Responsible Science: Ensuring the Integrity of the Research Process: Volume I. National Academy of Sciences (US), National Academy of Engineering (US), and Institute of Medicine (US) Panel on Scientific Responsibility and the Conduct of Research (National Academies Press, Washington, DC, 1992).

[29] L. K. John, G. Loewenstein, and D. Prele, Measuring the prevalence of questionable research practices with incentives for truth telling, Psychol. Sci. 23, 524 (2012).

[30] K. Fiedler and N. Schwarz, Questionable research practices revisited, Soc. Psychol. Pers. Sci. 7, 45 (2015).

[31] H. Fraser, T. Parker, S. Nakagawa, A. Barnett, and F. Fidler, Questionable research practices in ecology and evolution, PLoS ONE 13, e0200303 (2018).

[32] F. Agnoli, J. M. Wicherts, C. L. S. Veldkamp, P. Albiero, and R. Cubelli, Questionable research practices among Italian research psychologists, PLoS ONE 12, e0172792 (2017).

[33] A. M. Bogdan and A. F. Heckler, Effects of belief bias on student reasoning from data tables, in *Proceedings of the 2013 Physics Education Research Conference, Portland, OR*, edited by P. Engelhardt, A. Churukian, and D. Jones (AIP, New York, 2013), pp. 73–76.

[34] M. M. Stein, E. M. Smith, and N. G. Holmes, Confirming what we already know, in *Proceedings of the 2018 Physics Education Research Conference, Washington, DC*, edited by A. Traxler, Y. Cao, and S. Wolf (AIP, New York, 2018).

[35] E. M. Smith, M. M. Stein, and N. G. Holmes, Surprise! Shifting students away from model-verifying frames in physics labs, in *Proceedings of the 2018 Physics Education Research Conference, Washington, DC*, edited by A. Traxler, Y. Cao, and S. Wolf (AIP, New York, 2018).

[36] N. G. Holmes and D. A. Bonn, Quantitative comparisons to promote inquiry in the introductory physics lab, Phys. Teach. 53, 352 (2015).

[37] C. R. Gette, M. Kryjevskaia, M. R. Stetzer, and P. R. L. Heron, Probing student reasoning approaches through the lens of dual-process theories: A case study in buoyancy, Phys. Rev. Phys. Educ. Res. 14, 010113 (2018).

[38] Physics majors include students intending to major in physics or engineering physics.

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

[40] D. L. Schwartz and T. Martin, Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction, Cognit. Instr. **22**, 129 (2004).

[41] All names are pseudonyms.

[42] R. Khishfe and F. Abd-El-Khalick, Influence of explicit and reflective versus implicit inquiry-oriented instruction on sixth graders' views of nature of science, J. Res. Sci. Teach. **39**, 551 (2002).

[43] V. Dini, Investigating learners' epistemological framings of quantum mechanics, Ph.D. thesis, Tufts University, 2017.

[44] D. Overbye, Gravitational waves detected, confirming Einstein's theory, The New York Times, 2016, www .nytimes.com/2016/02/12/science/ligo-gravitational-waves-black-holes-einstein.html.

[45] M. Casey, *Scientists find evidence of gravitational waves predicted by Einstein* (Fox News, 2016), www.foxnews .com/science/scientists-find-evidence-of-gravitational-waves-predicted-by-einstein.

[46] T. Leopold, *Gravitational waves detected—and that's creating waves in science* (CNN, 2016), www.cnn.com/ 2016/02/11/us/gravitational-waves-feat/index.html.

[47] M. Wall, *Historic first images of a black hole show Einstein was right (again)* (Space.com, 2019), www .space.com/black-hole-event-horizon-images-einstein.html.

[48] K. Pimbblet, *Black hole image confirms Einstein's relativity theory* (EarthSky, 2019), www.earthsky.org/space/ black-hole-image-confirm-einsteins-theory-of-relativity.

[49] K. Pimbblet, *First black hole photo confirms Einstein's theory of relativity* (Phys.Org, 2019), www.phys.org/news/ 2019-04-black-hole-photo-einstein-theory.html.

[50] K. Houser, *The world's first black hole image vindicates Einstein's theory* (Futurism, 2019), www.futurism.com/ the-byte/black-hole-image-einstein-right-relativity.

[51] K. Pimbblet, *Scientists just took the first ever image of a black hole. To cut a long story short: Einstein was right* (NewsWeek, 2019), www.newsweek.com/first-black-hole-image-einstein-right-1393284.

[52] H. Ellis-Petersen, *Katie Bouman: The 29-year-old whose work led to first black hole photo* (The Guardian, 2019), https://www.theguardian.com/science/2019/apr/11/katie-bouman-black-hole-photo.

[53] We have reacted very emotionally to many aspects of this current research project.

[54] S. Allie, A. Buffler, L. Kaunda, and M. Inglis, Writing-intensive physics laboratory reports: Tasks and assessment, Phys. Teach. **35**, 399 (1997).

[55] D. Marasco, Teaching an old ball new tricks: Another look at energetics, motion detectors, and a bouncing rubber ball, Phys. Teach. **58**, 62 (2020).

[56] P. A. Bartlett and K. Dunnett, Secret objectives: Promoting inquiry and tackling preconceptions in teaching laboratories, arXiv:1905.07267.

[57] P. Bohacek, M. Vonk, J. Dill, and E. Boehm, Letting students discover the power, and the limits, of simple models: Coulomb's law, Phys. Teach. **55**, 380 (2017).

[58] M. Vonk, P. Bohacek, C. Militello, and E. Iverson, Developing model-making and model-breaking skills using direct measurement video-based activities, Phys. Rev. Phys. Educ. Res. **13**, 020106 (2017).

[59] B. 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).

[60] National Academy of Engineering, Practical Guidance on Science and Engineering Ethics Education for Instructors and Administrators, in *Papers and Summary from a Workshop December 12, 2012* (National Academies Press, Washington, DC, 2013), https://doi.org/10.17226/18519.

[61] National Research Council, *Research in the Life Sciences with Dual Use Potential: An International Faculty Development Project on Education About the Responsible Conduct of Science* (National Academies Press, Washington, DC, 2011), https://doi.org/10.17226/13270.

[62] National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, *On Being a Scientist: A Guide to Responsible Conduct in Research*, 3rd ed. (National Academies Press, Washington, DC, 2009), https://doi.org/10.17226/12192.

[63] National Academy of Engineering, Ethics education and scientific and engineering research, in *What's Been Learned? What Should Be Done? Summary of a Workshop* (National Academies Press, Washington, DC, 2009), https://doi.org/10.17226/12695.

[64] M. Kalichman, Responsible conduct of research (What, why, and does it work?), Acad. Med. **91**, e10 (2016).

[65] M. Kalichman, Rescuing responsible conduct of research (RCR) education, Account. Res. **21**, 68 (2014).

[66] A. L. Antes, S. T. Murphy, E. P. Waples, M. D. Mumford, R. P. Brown, S. Connelly, and L. D. Devenport, A meta-analysis of ethics instruction effectiveness in the sciences, Ethics Behav. **19**, 379402 (2009).

[67] D. L. Schwartz and J. D. Bransford, A time for telling, Cognit. Instr. **16**, 475 (1998).

[68] E. Etkina and G. Planinšič, Light-emitting diodes: Exploration of underlying physics, Phys. Teach. **52**, 212 (2014).