Students' views of experimental physics after participating in lab courses with open-ended activities

Qiaoyi Liu [*] and H. J. Lewandowski

JILA, National Institute of Standards and Technology and University of Colorado, Boulder, Colorado 80309, USA
and Department of Physics, University of Colorado, Boulder, Colorado 80309, USA

(Received 22 July 2025; accepted 12 November 2025; published 5 December 2025)

Improving students' views of experimental physics is often an important goal of undergraduate physics laboratory courses. However, traditional lab courses typically include highly guided activities that often do not require or encourage students to engage in the authentic process of experimental physics. Alternatively, open-ended activities in lab courses can provide students with a more authentic learning experience. Here, we investigate the correlation of open-ended activities in lab courses with students' views of experimental physics, using an expanded database of the Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS) survey responses from over 30,000 students at over 100 institutions. We show that the inclusion of some open-ended activities is associated with more expertlike postinstruction responses relative to the courses that include only traditional guided activities, and the effect is larger for students with low preinstruction scores. We also find that for courses with at least 1 week spent on open-ended activities, the number of weeks (above zero) spent on open-ended activities is not associated with a pre- to post-instruction gain in the E-CLASS scores. This result suggests that possibly including even a small amount of time on open-ended activities can improve students' views of experimental physics, which lowers the barrier for instructors to implement such components. These results also suggest that additional research should be conducted to better understand and categorize lab environments beyond simple definitions of inquiry, so as to understand the possible causation mechanisms of labs on students' views of experimental physics.

DOI: 10.1103/5np1-mccl

I. INTRODUCTION

Physics laboratory courses are generally considered to be an important component of the undergraduate curriculum [1,2]. These courses can provide students with valuable opportunities to engage in authentic scientific practices, develop practical lab skills, and work collaboratively with other students. Creating such opportunities requires investments in physical space, laboratory equipment, and instructor support. With the abundant opportunities and resources in many lab courses comes a myriad of possible, and often disparate, learning goals for these courses. Given the wide range of potential goals and limited class time, concerns are frequently raised about how effective such courses are at fulfilling the goals that they choose to focus on.

One of the concerns arises from the gap between the nature of practices present in many of the lab classrooms and those in professional scientific research and engineering labs. More specifically, a large portion of undergraduate physics lab courses are currently taught using only traditional guided lab activities, often dubbed "cookbook" labs. For these guided activities, the topics and methods of investigation are predetermined by the instructors, often providing specific directions so that students seek to confirm known outcomes [3]. These lab activities tend to fall short of replicating the cognitive tasks that commonly occur in authentic physics experiments [4] and thus potentially hinder the development of expertlike views in experimental physics for students [5].

In response to the critiques of traditional guided lab activities, members of the physics education community have worked on transforming courses at their institutions to allow students to engage in the process of experimental physics in a more authentic fashion [6–8]. A key feature of these course transformations is the inclusion of open-ended lab activities, where students are provided with more opportunities in making decisions regarding the selection of interesting phenomena to investigate, the design of

[*]Contact author: qili7958@colorado.edu

experimental apparatus, the choice of analysis methods, etc., rather than being asked to follow specific, predetermined directions in a lab guide. Additionally, a number of branded pedagogical approaches to lab instruction that are widely used in many institutions also incorporate open-ended hands-on activities. Examples of these pedagogical approaches include the Investigative Science Learning Environment (ISLE) [9], Modeling Instruction [10], Student-Centered Activities for Large Enrollment University Physics (SCALE-UP) [11], Thinking Critically in Physics Labs [12], and Course-based Undergraduate Research Experience (CURE) [13–16]. Although previous literature presents how effective some of these pedagogical approaches are in terms of achieving their respective learning goals, there is still much left unstudied when it comes to quantifying how open-ended activities in lab courses can improve students' views of experimental physics more generally, which is a common learning goal for most lab courses.

To measure students' views of experimental physics before and after lab instruction on a large scale, a lab-focused assessment was developed known as the Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS) [6,17,18]. E-CLASS is a research-based assessment tool that targets students' epistemologies and expectations about experimental physics, as well as student affect and confidence when performing physics experiments. It was validated through student interviews and expert review [19], and was tested for statistical validity and reliability using responses from students at multiple institutions and at multiple course levels [20].

A previous study compared thousands of students' E-CLASS scores and found that the inclusion of some open-ended lab activities in a lab course correlates with more expertlike responses after instruction as compared to courses that include only traditional lab activities [5]. However, the extent to which students' views of experimental physics can be improved by a certain amount of open-ended activities included in the lab course was not explored in this previous work, as it aggregated all courses that use any open-ended activities together in an effort to preserve statistical power in the analysis. Inspired by this study [5], we investigated how the *number of weeks* of open-ended activities correlates with the overall E-CLASS scores [21]. Determining how much of a lab course should be devoted to open-ended activities to achieve gains on the E-CLASS is important information for instructors wishing to focus on this goal, as these types of activities can be more challenging to implement in many environments.

Building on the prior works [5,21], this study utilizes a significantly larger number of student responses to the E-CLASS survey, which allows us to take a more detailed look at how open-ended activities impact E-CLASS scores. Unlike previous studies, we are able to probe how the number of weeks of open-ended activities correlates with not only the overall E-CLASS scores but also the item-by-item E-CLASS scores. Here, we aim to answer the following two research questions:

RQ1: To what extent do open-ended activities in lab courses correlate with the students' overall E-CLASS scores and individual E-CLASS item scores post-instruction as compared to courses with only guided labs? Does the correlation depend on their initial E-CLASS score before the course?

RQ2: How many open-ended activities, as measured by the number of weeks in a lab course, are necessary to see an improvement in students' overall E-CLASS scores and individual E-CLASS item scores?

Understanding how much open-ended activities are needed to see positive results could serve as a guide for lab course instructors to incorporate a sufficient amount of open-ended activities that could significantly improve their students' views of experimental physics. Moreover, the item-by-item analysis could provide valuable insights into the impact of open-ended activities on specific E-CLASS items, so that instructors are informed on how open-ended activities are related to a certain aspect of experimental physics and adjust their course structure and pedagogy accordingly in future course planning.

We begin by providing relevant background information on lab pedagogy in the realm of traditional guided and open-ended lab activities, as well as prior education research findings related to students' views of experimental physics using E-CLASS (Sec. II). This is followed by a description of the methodology used in our investigation (Sec. III). We explain the procedure for data collection and cleaning, then outline the statistical analysis techniques employed to analyze the E-CLASS data, and finally highlight the limitations of this study. The results and discussion (Sec. IV) presents the findings of the study, as well as provides an interpretation of the findings and their alignment with the research questions. Finally, the conclusion (Sec. V) summarizes the key findings of the study, reiterating the impact of open-ended activities on students' views of experimental physics. We conclude by exploring a variety of options for future work that could be conducted to enrich our understanding of open-ended lab activities beyond this study.

II. BACKGROUND

In this section, we provide an overview of the background of existing lab pedagogy in the area of traditional guided and open-ended lab activities. Additionally, we review previous studies that have utilized E-CLASS as an assessment tool and highlight the main findings they have yielded. We specifically discuss Wilcox and Lewandowski [5], as it is particularly relevant to the topic explored in this study.

A. Existing lab pedagogy

Traditionally, physics lab courses have been taught with little student choice in the scientific questions, procedure, or even experiment they do. In these traditional guided lab courses, students typically follow highly structured, step-by-step instructions, on what to do and how to do it: the relevant equations and principles are often laid out in advance; students are provided with the equations to predict particular measurements or even directly informed of the exact values they are expected to obtain; at times they are even directed as to what exact actions to take with the equipment to produce the "desired" output [3,4]. This lack of decision making does not require students to do much of the critical thinking and reasoning that consistently occurs in authentic physics experiments [3,4].

As a result, there are many calls to transform physics lab courses from the physics education community [3,22], and national science recommendations for the lab curriculum [23,24]. One theme that spans these calls is that instructional labs should move away from traditional guided teaching methods in favor of open-ended ones that support more student decision making. To make independent and substantial decisions, students should be able to enact their agency, the capacity to guide one's action intentionally toward achieving a goal within certain constraints [25,26]. When conducting an experiment, physicists face various constraints in their projects, such as time and space restrictions, as well as budget allowance. Within these restrictions, however, they can make many independent and deliberate decisions when carrying out their experiments. Therefore, promoting student agency in physics lab courses can be relevant in replicating the experience of authentic scientific practices, where students can have the opportunity to investigate their own research questions, make deliberate decisions in their experimental designs, or choose appropriate analysis methods and tools, etc. Open-ended lab activities that support student agency can have students more actively involved in their learning process, so they can construct new knowledge through the authentic scientific process. When given decision-making power in lab instruction, students are shown to be more engaged [27], develop confidence [28], ownership [29], and increase engagement in expertlike experimentation behaviors [26,30].

Despite the many proven benefits of open-ended lab instruction for various student outcomes, there can be some challenges creating student buy-in during its implementation. For instance, many first-year students may not have experienced the open-ended lab format in high school, because either they have only experienced traditional guided physics labs or have never taken a physics lab course before entering college [31]. In turn, when entering an open-ended introductory physics lab, these students might not be familiar with the open-ended format, and thus may develop resistance to this sudden shift in paradigm. More specifically, these students may feel less comfortable with the lack of structure and guidance [32] and prefer more authoritative instruction [33]. To address this challenge of student buy-in, Kalender *et al.* suggested instructors gradually reduce the structure in the lab course by incrementally providing students with more opportunities to make their own decisions [34]. A course transformation following this suggestion started with a more guided inquiry format and transitioned into more open-ended inquiry as the semester progressed, and research has shown that most of the students enrolled in this course perceived the opportunities of agency positively and irrespective of student gender [34].

In similar efforts to support student agency in instructional labs, many institutions have implemented course transformations that incorporated varying degrees of open-ended activities, allowing students to make decisions in their investigations. These diverse approaches to lab instruction have led to a range of student outcomes. Doucette *et al.* documented the transformation of a beyond-first-year lab course, which was implemented in three lab modules spanning the course of a semester and focused on student engagement in troubleshooting by introducing new technology [7]. The technological aspects of the transformation were well received by a majority of the students, but some struggled with the limited amount of scaffolding in the course [7]. Hoehn *et al.* presented a case study analysis of one advanced lab course that transitioned from traditional labs to more open-ended, four-week-long projects with the transition to remote teaching due to the COVID-19 pandemic [8]. They reported that the students responded positively about the open-ended nature of the projects, but some reported a high workload and unclear expectations made the course overwhelming and stressful [8].

Additionally, a number of branded pedagogical approaches to lab instruction that are widely used in many institutions also incorporate open-ended hands-on activities to promote student agency in different ways. For instance, ISLE encourages the use of authentic scientific practices while learning physics, with a strong emphasis on students expressing and exploring their own ideas without being asked to make predictions [9]. In the ISLE approach, students begin by observing natural phenomena and identifying patterns, then develop explanations for these patterns and use them to make predictions. Modeling Instruction has a lab component focused on constructing physics knowledge by developing and deploying mathematical models [10]. SCALE-UP also has a lab component that emphasizes designing experiments and analyzing data [11]. The Thinking Critically in Physics Labs aims to teach students about the nature of scientific experimentation and to develop their experimentation and critical thinking skills [12]. Finally, a growing trend is to transform traditional guided lab courses into CUREs. In CUREs, while engaging in authentic scientific practices, students address research questions with an outcome that is unknown both to the students and to the instructors [13,14].

B. E-CLASS

Across the wide range of modifications to the physics lab curriculum, a common learning goal is to shift students' views of experimental physics to better align with those of professional experimental physicists. This shared goal creates a need for evaluation tools that help instructors improve their courses and provide researchers with insights into how different course modifications affect student learning. With the aim of fulfilling this need, E-CLASS was specially designed to assess the impact of lab courses on students' views of experimental physics at all stages of an undergraduate program [6,17,18]. In the context of a lab course, students' views of experimental physics refer to defining what is viewed as a good or valid experiment and what are the appropriate ways to understand the design and operation of an experiment and the communication of results [35]. E-CLASS measures how students personally perceive the strategies, habits of mind, and attitudes toward experimental physics while conducting experiments in lab classes, as well as how they predict those same strategies, habits of mind, and attitudes are practiced by expert physicists [19,35]. E-CLASS is well suited to assess some dimensions of lab instruction for at least three reasons. First, it is not directly tied to specific physics content, which increases its applicability in the preexisting wide range of lab courses. Second, the habits of mind and ways of thinking probed by E-CLASS represent an important and common learning goal for experimental lab courses for many instructors and national calls. Third, E-CLASS can be valuable when evaluating educational environments that have significant differences from professional practice, so that instructors can identify areas where improvements and adjustments are needed to bridge the gap and better prepare students through their physics lab courses.

E-CLASS presents students with a total of 30 statements related to various aspects of experimental physics, including affect, argumentation, confidence, experimental design, math-physics-data connections, modeling the measurement system, physics community, purpose of labs, statistical uncertainty, systematic error, and troubleshooting, and asks them paired Likert-style questions about their own perceptions of doing experiments in class, as well as how they perceive an experimental physicist might respond when working on their own research [19]. It is crucial to note that the categories listed above should not be misconstrued with latent variables associated with E-CLASS, where one expects a high degree of correlation among the items within a category. For instance, the two statements in the physics community category are "scientific journal articles are helpful for answering my own questions and designing experiments" and "communicating scientific results to peers is a valuable part of doing physics experiments." Both statements express how experimental physics extends beyond the individual researcher, but they clearly represent distinct practices, and any particular lab course could emphasize them to varying degrees [19]. In fact, E-CLASS does not exhibit a strong factor structure (which is aligned with its original design) according to previous research that conducted factor analyses on E-CLASS data [20,36]. Additionally, the survey developers have cautioned using the overall E-CLASS scores to interpret the results for a specific course, as all 30 statements might not be connected to the goals of every course [5,20]. Regardless, the overall E-CLASS is still useful in that it provides a continuous variable that offers a holistic view of students' performance on E-CLASS [5].

The current version of E-CLASS has now been administered as an online pre- and post-test format since 2012 at multiple institutions across the United States and across the world, and the data collection is still ongoing. Between 2016 and 2019, a total of 22,541 matched pre- and post-responses to the E-CLASS survey were collected, covering 133 institutions, 599 unique courses, and 204 instructors [36]. This large dataset served as an example of sharing quantitative data in the field of physics education research through open access [36], and has been used to characterize the landscape of physics lab instruction across the United States [37].

Several previous studies have used E-CLASS to understand the impact of specific course transformations on students' views of experimental physics at a single institution. For instance, Doucette *et al.* found that students enrolled in a conceptual inquiry-based lab curriculum supplemented with reflection questions were associated with higher postinstruction E-CLASS scores as compared to those in traditional physics curriculum at University of Pittsburgh [38]. Sulaiman *et al.* studied more than 3000 student responses collected before and after a transformation focused on developing experimental skills and expert-like epistemology in an introductory physics lab course at University of Colorado, examining overall E-CLASS scores, as well as item-by-item scores. The results indicated a statistically significant increase in the overall E-CLASS score after the transformation, irrespective of student gender [39]. Moreover, the item-by-item analysis revealed larger gains in a few E-CLASS items, especially those associated with the new course learning goals [39]. Werth *et al.* analyzed student responses collected before and after the Colorado Physics Laboratory Academic Research Effort (C-PhLARE) CURE at University of Colorado [40]. The item-by-item analysis showed that students enrolled in the C-PhLARE CURE were associated with significantly higher scores on many of the E-CLASS items, especially those that aligned with the objectives of the course, as compared to historical, open-source E-CLASS dataset from first-year physics lab courses [36,40].

Additionally, many previous studies have used the aggregate E-CLASS dataset to explore the correlation between various general instructional conditions and students' views of experimental physics at multiple institutions. For instance, Fox *et al*. analyzed E-CLASS responses from over 1600 students in both spring and fall semesters of 2020 during the emergency remote instruction due to the COVID-19 pandemic, and found that students in 2020 had similar overall E-CLASS scores compared to those in 2019, indicating that there was no net effect on students' views of experimental physics due to the remote modality of the lab [41]. Furthermore, a series of studies by Wilcox and Lewandowski explored the influence of various instructional factors on students' E-CLASS scores [5,20,42–46]. A summary of their results can be found in Ref. [18]. In this series of studies, they revealed that E-CLASS postinstruction scores have little to no correlation with students' final course grade [42], that courses using research-based curricula (including ISLE, Modeling Instruction, and SCALE-UP) realize higher postinstruction E-CLASS scores on average than those using only traditional guided labs [43], that courses focused on developing lab skills demonstrated higher postinstruction E-CLASS scores than those that concentrated on reinforcing physics concepts or pursuing both goals [45], and so on. The last finding was later confirmed by Walsh *et al*. using a more expansive database of survey responses from more than 20,000 students across 100 different institutions [47].

One of the most important studies in this series is the one that investigated the impact of open-ended versus guided lab activities on students' views of experimental physics. Wilcox and Lewandowski analyzed a total of 4915 matched student responses to the E-CLASS survey from 147 courses collected between Spring 2015 and Spring 2016 [5]. Even though the relative fraction a particular course spent on open-ended activities varied significantly, 84 out of the 147 courses reported having at least 1 week of open-ended activities, which include 1149 out of the 4915 students. To preserve statistical power, all courses using any open-ended activities were aggregated together as a single group, and the entirety of the analysis treated courses dichotomously as either having open-ended activities or having only traditional guided lab activities. First examining the differences in the raw pre- and post-instruction E-CLASS scores, they found that students' overall E-CLASS scores decreased after traditional guided labs and did not decrease after open-ended labs. Moving on to using an analysis of covariance (ANCOVA), they discovered that courses using open-ended activities were associated with higher postinstruction E-CLASS scores compared to those using only traditional guided activities, and this correlation persisted even after controlling for factors such as preinstruction scores, course level, student major, and student gender. However, a key limitation of this study is that the limited scope of the E-CLASS dataset at the time prevented exploration of how varying amounts of open-ended activities in the lab course influenced students' views on experimental physics. Now that significantly more E-CLASS responses have been collected since the previous work, we have revisited this topic and performed a more fine-grained analysis of how the number of weeks of open-ended activities correlates with the overall E-CLASS scores [21]. We found that the inclusion of some open-ended activities is associated with more expertlike postinstruction responses relative to the courses that include only traditional guided activities, and that the number of weeks spent on open-ended activities is not associated with pre- to post-instruction gain in E-CLASS scores. We expand upon this work now using a more complete and recent version of the E-CLASS dataset, and reexamine how the number of weeks of open-ended activities correlates with not only the overall E-CLASS scores but also the item-by-item E-CLASS scores.

III. METHODS AND DESIGN

In this section, we provide an overview of what E-CLASS is and how it is scored, describe the data source and student demographics, present the statistical methods used to address the two research questions, and point out the limitations of our methodology that may affect the interpretation of this study.

A. E-CLASS

We used E-CLASS to quantitatively measure the impact of open-ended activities in lab courses on students' views of experimental physics. Students indicate their level of agreement with each of the 30 statements on a five-point Likert scale from "strongly agree" to "strongly disagree" based on two questions: "What do YOU think when doing experiments for class?" and "What would experimental physicists say about their research?" For this study, we used responses to only the first question. Additionally, students have the option to provide additional demographic information, such as gender, race or ethnicity, and major. It is worth noting that student race or ethnicity information was not collected prior to 2016 [20].

For scoring purposes, students' responses to each five-point E-CLASS item were condensed into a standardized, three-point scale in which responses "(dis)agree" and "strongly (dis)agree" were collapsed into a single (dis) agree category. Students' responses to individual items were scored based on consistency with the expert response: $+1$ for a response consistent with the experts, 0 for neutral, and -1 for a response inconsistent with the experts [20]. The collapse of the five-point scale to three points is common in the analysis of Likert-style items, and is partially motivated by the inherently ordinal, rather than interval, nature of the Likert scale [48]. The use of a three-point scale is also supported by previous literature suggesting that the threshold between "(dis)agree" and "strongly (dis)agree" is not always consistent between individual students or groups with different cultural backgrounds [49]. Additional discussion of different scoring schemes can be found in Ref. [20]. A student's overall E-CLASS score is then given by the sum of their scores on each of the 30 items, resulting in a possible score range of

[−30, 30]. We can also calculate the average score per item for a group of students, resulting in an average in the range of [−1, 1].

The E-CLASS was administered via an online system at the beginning (pre) and the end (post) of a course to measure the change in students' views of experimental physics [50]. To use E-CLASS through its centralized administration system, interested instructors first complete the Course Information Survey (CIS), which collects logistical, structural, and pedagogical information about the course. Using the information from the CIS, the system then generates unique links to the pre- and post-instruction E-CLASS for each course and sends them to the instructors to distribute to their students at the beginning and the end of the course. After students have completed the surveys, the system stores the raw data on a local server and subsequently allows for the aggregation of a large-scale dataset of students' responses.

B. Data source

The data used for this study were collected from undergraduate physics lab courses between Spring 2015 and Spring 2024 using the E-CLASS centralized administration system [50]. During this period, we collected a total of 70,967 preinstruction responses and 55,873 post-instruction responses. Only students for whom we had matched pre- and post-instruction responses were included in the analysis. The raw data were subsequently deidentified according to the established data sharing model, where each student randomly received a student ID number [36]. Pre- to post-instruction matching was conducted based on the student ID number. The E-CLASS also includes a filtering question to eliminate responses from students who did not read the item prompts, so that any student who responded incorrectly to this filtering question was also removed from the analysis [20,36]. The E-CLASS matched and filtered dataset used in this study includes a total of 40,594 matched student responses, which is about an order of magnitude more data than the previous study on this topic.

Over the nine years of data collection, we aggregated instructor responses to the CIS from 1099 distinct instances of courses. Out of these 1099 courses, 170 of these did not have students fill out E-CLASS both pre- and post-instruction, and so we have matched pre- and post-instruction E-CLASS student responses from 929 of these courses. In the CIS, instructors were asked to report how many weeks of instruction were spent on "all guided lab activities" and how many weeks were spent on "all open-ended activities or projects," and 798 out of the 929 courses reported this information. As a check, we summed the reported number of weeks spent on "all guided lab activities" and "all open-ended activities or projects" for each of the 798 courses, and decided that the 75 courses that reported more than 16 weeks of total instruction may interfere with the analysis, as they are likely year-long courses, or simply reporting errors, and thus these courses were removed from the final dataset. This resulted in a total of 30,566 matched student responses in 723 distinct courses spanning 157 institutions. These institutions included a variety of different types, with 53 Ph.D. granting institutions, 12 master's granting institutions, 81 4-year colleges, and 11 2-year colleges. Moreover, these courses span the full space of introductory and advanced labs, with 431 first-year and 292 beyond-first-year courses. It is

TABLE I. Institution types and course levels that students were enrolled in, as well as their demographic information in the final matched dataset.

	Total	Guided-only activities	Some open-ended activities
Total	30,566	22,304	8262
Institution type			
Ph.D. granting institution	24,101	19,713	4388
Master's granting institution	1716	1038	678
4-year college	4215	1344	2871
2-year college	534	209	325
Course level			
First year	26,787	20,887	5900
Beyond first year	3779	1417	2362
Gender			
Man	16,071	11,410	4661
Woman	13,558	10,275	3282
Other	937	619	318
Major			
Physics, engineering physics, astrophysics	4077	1980	2097
Other engineering	7186	5455	1731
Other science and math	14,931	11,583	3348
Nonscience majors	2789	2136	653
Open option, undeclared	1583	1150	433

important to note that the analysis we conducted is at the student level, and not institution or course level, and so we break down the number of students in the final matched dataset that are enrolled in different institution types and course levels, as well as their demographic information in Table I. Table I does not include a breakdown of the racial demographics of the students, because this information was not collected prior to 2016, which covers 3895 out of the 30,566 matched student responses. It does, however, report the breakdown of students by gender and major in the final matched dataset. Here, astronomy and computer science are included in the "Other science and math" category.

Out of the 723 courses, 367 of them had only guided lab activities with a total of 22,304 students enrolled, while 356 of them included at least 1 week of open-ended activities with a total of 8262 students enrolled. The distributions of the weeks spent on open-ended activities by course and by students enrolled are shown in Fig. 1. While the number of weeks spent on open-ended activities varied significantly among the courses and students, the overall distribution is skewed toward a lower number of weeks. The average course includes 4.9 ± 0.2 weeks of open-ended activities, and the average student experiences 5.82 ± 0.04 weeks of open-ended activities in a course. For informational purposes only, the distributions of fraction of weeks spent on

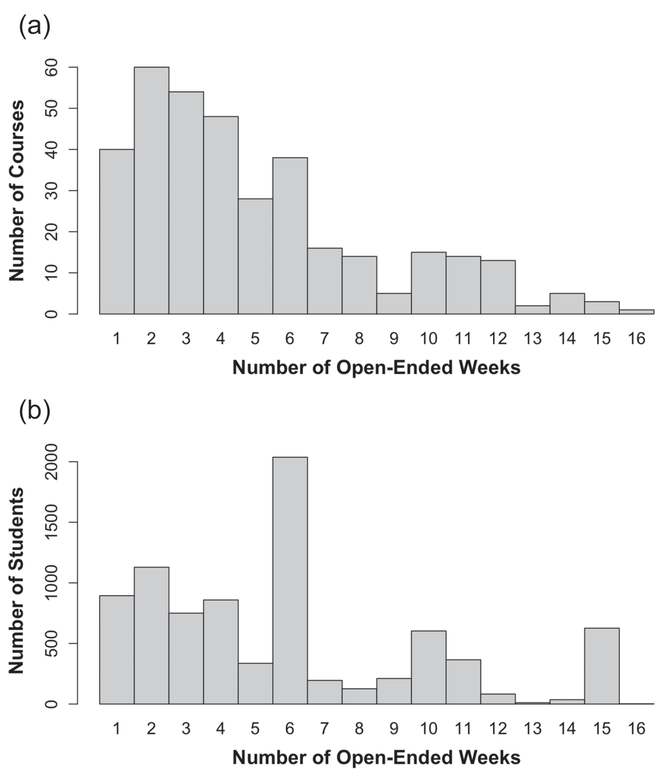

FIG. 1. Histograms of the number of weeks spent on open-ended activities for (a) the 356 lab courses that include at least 1 week of open-ended activities and (b) the 8262 students that are enrolled in these courses.

open-ended activities by course and students enrolled are presented in the Appendix A.

C. Analysis methods

To address one component of our first research question, regarding the impact of open-ended activities in courses as compared to fully guided ones, we treated the 723 courses dichotomously as either having open-ended activities, regardless of the number of weeks spent on those activities, or having only guided activities. We first examine the overall behavior by comparing the means of the E-CLASS scores pre- and post-instruction for students in courses using open-ended activities and those using only guided activities. As the distribution of scores on E-CLASS is typically skewed toward positive scores [20], we report statistical significance based on the nonparametric Wilcoxon signed-rank test unless otherwise stated [51]. We also report the Cohen's d along with its corresponding 95% confidence interval (CI) as measures of effect size and practical significance [52–54]. To explore this effect further, we utilize an analysis of covariance (ANCOVA) in addition to examining students' raw pre- and post-instruction E-CLASS scores [55]. ANCOVA is a statistical method for comparing the difference between population means after adjusting them to account for the variance associated with other variables. In this case, we want to determine whether the difference between the E-CLASS scores of students in courses using different types of lab activities, i.e., open-ended vs guided-only activities, remains statistically significant after accounting for differences in preinstruction scores.

To address one component of our second research question, regarding the amount of open-ended activities that is correlated with a significant change in students' views of experimental physics, we constrained our analysis to a subset of the full dataset where we include only the 356 courses with at least 1 week of open-ended activities, and exclude the other 367 courses with only guided activities. We chose the number of weeks spent on open-ended activities as the measure of the amount of open-ended activities included during the course. The rationale behind choosing the number of weeks spent on open-ended activities as the variable of interest, as opposed to fraction of weeks on open-ended activities, is that the total number of weeks is not the same for all the courses. A prime example of this is that courses in institutions on the semester system are typically longer than those in institutions on the quarter system. Due to the relatively continuous nature of the number of weeks, as opposed to the categorical nature of the instructional condition between open-ended and guided-only, we utilized a nested set of linear regression models to control for students' preinstruction scores.

Another component of our two research questions requires analysis of student performance on individual

E-CLASS items, as opposed to the overall E-CLASS score, and how this correlates with open-ended activities in lab courses. To do so, we employed ordinal logistic regression models [56], and performed predictive modeling instead of causal modeling. Ordinal logistic regression is a statistical method for predicting an ordinal response variable based on one or more explanatory variables. In this case, we want to predict the likelihood of students' agreement for each individual E-CLASS item postinstruction based on the types of lab activities, i.e., open-ended vs guided-only, as well as the amount of open-ended activities, i.e., the number of weeks spent on open-ended activities, after accounting for the influence of preinstruction scores.

To assess the significance of the model coefficient, we computed the p value using the t distribution. We adopted a significance level α of 0.05, a common p value threshold for statistical significance. However, given that we are simultaneously running these models on all 30 individual E-CLASS items, we may run into the multiple comparison problem. The problem with multiple comparisons is that as the number of comparisons on the same sample increases, it becomes more likely to obtain a significant result for at least one comparison, purely due to chance, even if the groups being compared are uncorrelated [57,58]. To counteract the multiple comparison problem, we employed the Holm-Bonferroni method, which adjusts the p values to account for the number of statistical tests [57,58]. In this case, a Holm-Bonferroni corrected p value, denoted here as $\tilde{p}$, less than the significance level α of 0.05 indicates a statistically significant relationship between the predictor variable (instruction type or number of weeks spent on open-ended activities) and the odds of a more expertlike response on an individual E-CLASS item.

We also calculated odds ratios (OR) by exponentiating the coefficient estimates of the model. Odds ratios represent the multiplicative change in the odds of moving up a response category (e.g., from inconsistent with experts to neutral, or from neutral to consistent with experts) associated with changing from one group to an adjacent group (e.g., from guided-only to open-ended, or 1 week to 2 weeks spent on open-ended activities). Therefore, the odds ratios with confidence intervals that overlap with 1 are not statistically significant. On the other hand, odds ratios with confidence intervals that are strictly greater than 1 suggest that there is an increased likelihood of having a more expertlike E-CLASS response by moving from one group to an adjacent group in a particular direction, based on the ordering of the groups, while those with confidence intervals that are strictly less than 1 suggest that there is an increased likelihood of having a less expertlike E-CLASS response when comparing those adjacent groups in the same direction. The analysis was conducted in both Python and R to ensure that the results were in agreement with the two statistical packages [59,60].

D. Limitations

We would like to acknowledge the limitations associated with the methodology, serving as a preface to readers when interpreting our results to follow.

First, we are using the number of weeks spent on open-ended activities as the proxy for the amount of open-ended activities included in a lab course, but it may not be the best option for all courses. The terms "guided lab activities" and "open-ended activities" were not defined in the CIS, and different instructors may view these characterizations differently.

Moreover, instructors for the courses in our dataset generally chose to use E-CLASS in their classrooms voluntarily and likely care about the items that E-CLASS poses, so these courses in our dataset are not randomly selected and not representative of all lab courses. The fact that we are including only courses whose instructors reported the information on the guided and open-ended activities may also contribute to this selection bias.

In addition, while the dataset we are studying is extensive, spanning a large number of institutions, courses, and student populations, the analysis is still limited by sample selection. For example, as can be seen in Table I, students enrolled in 2-year colleges are underrepresented in our dataset. Additionally, as can be seen in Fig. 1, there are only a few courses with a relatively large number of open-ended weeks. The data are not uniformly weighted among these variables, meaning some were more precisely measured than others, and in turn some results may be biased toward particular institution type or course feature.

Finally, this study is observational, and the purely quantitative analysis that we report speaks to what occurs among the students in these courses, but not why the outcomes occurred. Any statements made here that relate to causality are tentative and hypothetical and the results of this study should point to where specific research needs to be conducted in the future to understand causal relationships.

IV. RESULTS AND DISCUSSION

In this section, we present and discuss the findings by addressing each research question. The first examines differences in E-CLASS scores between courses that include only guided lab activities and those that incorporate any amount of open-ended activities. The second explores how E-CLASS scores vary with the number of weeks students engage in open-ended activities.

A. RQ1: Comparing courses with only guided activities to courses with open-ended activities

To explore general trends in the aggregate data, we first examine the differences in raw pre- and post-instruction E-CLASS scores for students in courses using open-ended activities and those using only guided activities, as shown in Table II. Students in courses using open-ended activities showed a small, but statistically significant positive shift from before to after instruction, while those in courses using only guided activities showed a relatively larger and statistically significant negative shift. Moreover, students in courses using open-ended activities score higher than those in courses using only guided activities, both preinstruction ($p = 5.06 \times 10^{-19}$) and postinstruction ($p = 2.34 \times 10^{-99}$). While the difference is statistically significant both before and after instruction, the magnitude of this effect is larger for the postinstruction scores ($d = 0.269$, 95% CI [0.256, 0.282]) than the preinstruction scores ($d = 0.116$, 95% CI [0.103, 0.129]). The comparison of E-CLASS scores between students in courses using different types of lab activities suggests that open-ended activities have a positive impact on students' views of experimental physics, as compared to the fully guided ones, which is consistent with the results shown in Wilcox and Lewandowski with a smaller dataset [5].

Since we observed a small, but statistically significant, difference in the preinstruction E-CLASS scores between the two conditions, we performed a one-way ANCOVA to compare postinstruction E-CLASS scores for courses using guided and open-ended activities, while using the preinstruction score as a covariate. In order for the results of an ANCOVA to be valid, the data must meet several assumptions, one of which is that the slope of the regression line between the dependent variable and covariate is the same for each category of the independent variable [61]. To verify this assumption for the E-CLASS matched data, we initially employed the following model, which included an interaction term between preinstruction score and the instruction type:

$$PostECLASSScore \sim PreECLASSScore + OpenEnded + PreECLASSScore \times OpenEnded. \tag{1}$$

In this model, $PreECLASSScore$ and $PostECLASSScore$ correspond to students' E-CLASS scores pre- and post-instruction, while the categorical variable $OpenEnded$ corresponds to the instruction type, and is coded as 0 for students in courses using only guided activities and 1 for those in courses using open-ended activities. After fitting this model to the E-CLASS matched data, we found that there is a statistically significant interaction between preinstruction score and the instruction type ($F = 68.4$, $p = 1.41 \times 10^{-16}$). The expected postinstruction E-CLASS scores from our model for students in courses using guided and open-ended activities according to the ANCOVA results are plotted in Fig. 2, which further confirms that the slopes are unequal between the two instruction types. This result means that open-ended activities have a positive effect on postinstruction scores for all students on average, but a larger positive effect for students with low preinstruction scores.

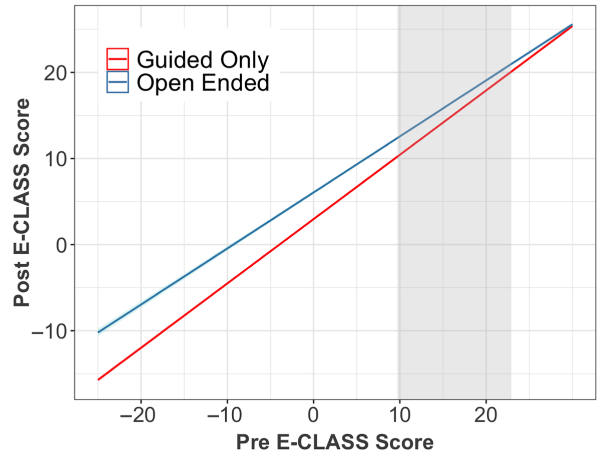

FIG. 2. Expected postinstruction E-CLASS scores as a function of preinstruction E-CLASS scores for students in courses using only guided activities (solid red line) and those using some open-ended activities (solid blue line). While the one standard deviation intervals are plotted here, they appear invisible with respect to the thickness of the linear regression lines due to the large sample sizes. The shaded gray region represents the range of preinstruction scores centered around the global mean with a width of twice the global standard deviation.

Tests of the E-CLASS matched data showed that they satisfied all the ANCOVA assumptions, except homogeneity of regression slopes, as is indicated by the statistically significant interaction term. We subsequently ran the one-way ANCOVA model without the interaction term, with this assumption violation in mind. After fitting this

TABLE II. Overall mean E-CLASS scores and their standard errors (SEs) on both the pre- and post-tests for students in courses using only guided activities and those using open-ended activities. Significance and Cohen's d, reported as 95% confidence interval (CI), describe the difference between students' scores in the guided-only and open-ended courses.

	Pre E-CLASS score (Mean ± SE)	Post E-CLASS score (Mean ± SE)	Wilcoxon signed-rank test p	Cohen's d [95% CI]
Guided-only activities	16.11 ± 0.04	15.00 ± 0.05	1.02×10^{-123}	−0.154 [−0.164, −0.145]
Some open-ended activities	16.87 ± 0.07	17.03 ± 0.08	6.01×10^{-7}	0.024 [0.008, 0.040]

new model to the E-CLASS matched data, we found that there is a statistically significant effect of the instruction type on the postinstruction score after controlling for preinstruction score ($F = 382$, $p = 1.61 \times 10^{-84}$), although the effect size is fairly small (partial $\eta^2 = 0.0123$). Since the homogeneity of slope assumption is violated, our results here should be interpreted as lower bounds for the partial η^2 instead of absolute values on the relationship between the instruction type and postinstruction E-CLASS score [62].

Finally, we can examine the difference between courses using open-ended activities and those using only traditional guided activities on an item-by-item level using an ordinal logistic regression model that predicts the likelihood of students' agreement for each individual E-CLASS item after instruction based on the instruction type, which uses the preinstruction score as a covariate and includes an interaction term between preinstruction score and the instruction type:

$$\log\left[\frac{\Pr(E_{post} \leq j)}{\Pr(E_{post} > j)}\right] = \alpha_j + \beta_{pre}E_{pre} + \beta_{open}OpenEnded + \beta_{int}E_{pre} \times OpenEnded \tag{2}$$

In this model, $\Pr(E_{post} \leq j)$ represents the cumulative probability that the postinstruction score on an individual E-CLASS item falls in response category j or lower (e.g., from "inconsistent with experts" to "neutral," or from "neutral" to "consistent with experts"). The complementary probability, $\Pr(E_{post} > j)$, reflects the likelihood of the score falling in a higher response category. The parameter α_j is the intercept for response category j. The coefficient β_{pre} corresponds to the effect of the preinstruction score for that E-CLASS item. The coefficient β_{open} captures the effect of the instruction type variable. Finally, β_{int} represents the interaction effect between the preinstruction score and instruction type.

After fitting this model to the E-CLASS matched dataset, we found that only two out of the 30 E-CLASS items have statistically significant interactions between preinstruction score and instruction type after the Holm-Bonferroni correction, namely, item 14 ($\tilde{p} = 0.000399$) and item 27 ($\tilde{p} = 0.0197$). (See the Appendix B for the E-CLASS item statements.) For the purpose of reporting odds ratios for the individual E-CLASS items, we opt to remove the interaction terms, and model the postinstruction score as a function of instruction type with the preinstruction score as a covariate, because odds ratios for the stand-alone terms in models with interaction terms do not have meaning due to the collinearity between the stand-alone and interaction terms. However, significant interaction terms indicate that when interpreting odds ratios, readers should take caution to remember that the full effect is not explained by these two variables alone for these two items, but rather the one impacts the other.

A list of all the individual E-CLASS items, their corresponding item numbers, along with their associated Holm-Bonferroni corrected p values, odds ratios, and 95% confidence intervals from the ordinal logistic regression are presented in the Appendix B and plotted in Fig. 3. As can be seen in Fig. 3, 24 out of the 30 E-CLASS items

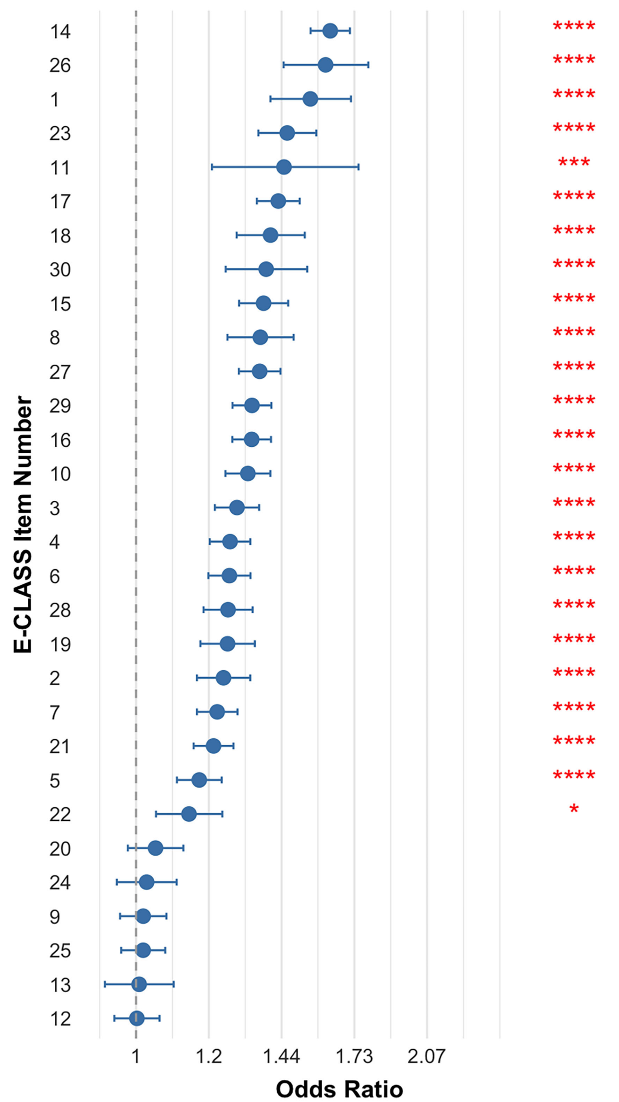

FIG. 3. Odds ratios and their corresponding 95% confidence intervals comparing students in courses with only guided activities and ones with some amount of open-ended activities. Odds ratios represent the multiplicative change in the odds of moving up a response category. An odds ratio greater than 1 (indicated by the vertical dashed line) signifies a higher likelihood of students in courses with open-ended activities having a more expertlike response in the post-test for the respective E-CLASS item compared to those in courses with only guided activities. Additionally, statistical significance levels are denoted next to each item. Significance at a Holm-Bonferroni corrected p value of less than 0.0001 is represented by ****, p less than 0.001 is ***, p less than 0.05 is *, and p values greater than 0.05 are not symbolized.

demonstrated statistically significant odds ratios after the Holm-Bonferroni correction, all with their associated 95% confidence intervals strictly greater than 1, with all in favor of courses with open-ended activities. This indicates that after controlling for the preinstruction scores, a student in a course with open-ended activities is more likely to respond to a vast majority of these E-CLASS items in a more expertlike fashion after instruction than a student in a course with only traditional guided activities. This positive impact of open-ended activities as compared to the guided ones indicates that open-ended activities have the potential to develop many aspects of expertlike views of experimental physics. In particular, the E-CLASS item with the highest odds ratio is item 14, which states that "when doing an experiment, I usually think up my own questions to investigate." Students in a course with open-ended activities were roughly 1.6 times more likely to develop a more expertlike views about developing their own research questions after instruction than those in a course with only traditional guided activities. This is directly associated with student agency in decision making in their lab courses.

Overall, the results of the descriptive statistics and ANCOVA on the overall E-CLASS scores indicate that students in courses with open-ended activities all scored significantly higher than those in courses with only traditional guided activities after instruction, while controlling for the preinstruction score, but the shift is more significant for students with low preinstruction scores, as shown in Fig. 2. This suggests that open-ended activities may have a greater benefit for students who started with less expertlike views of experimental physics, as measured by E-CLASS. Moreover, the results of the ordinal logistic regression indicate that a student in courses with open-ended activities is more likely to score significantly higher on many of the E-CLASS items than those in courses with only traditional guided activities after instruction, while controlling for the preinstruction score. This finding is consistent with previous research on the overall E-CLASS scores [5,21] (now using a much larger dataset), confirming the positive influence of open-ended activities on students' views of experimental physics, as compared to the traditional guided ones.

B. RQ2: Comparing E-CLASS scores for varying number of weeks spent on open-ended activities

To provide a more detailed picture of how much open-ended activities might improve students' views of experimental physics, we investigated a subset of the E-CLASS matched data that includes only courses with at least 1 week spent on open-ended activities. Similar to the analysis procedure for RQ1, we first examined the overall E-CLASS scores by fitting the following linear regression model to the data, which includes an interaction term between preinstruction score and the number of weeks spent on open-ended activities:

$$\begin{aligned} PostECLASSScore = \beta_0 &+ \beta_{pre} PreECLASSScore \\ &+ \beta_{open} OpenWeeks \\ &+ \beta_{int} PreECLASSScore \\ &\times OpenWeeks \end{aligned} \quad (3)$$

In this model, $OpenWeeks$ corresponds to the number of weeks spent on open-ended activities, β_0 is the intercept of the model, the coefficient β_{pre} represents the mean change in the postinstruction score per unit change in the preinstruction score, the coefficient β_{open} represents the mean change in the postinstruction score per unit change in the number of weeks spent on open-ended activities, and the coefficient β_{int} estimates the interaction effect between the preinstruction score and number of weeks spent on open-ended activities. After fitting this model to the subset of the E-CLASS matched data (excluding courses with $OpenWeeks = 0$), we found that there is a significant interaction between preinstruction score and the number of weeks spent on open-ended activities ($\beta_{int} = 0.006 \pm 0.003$, $p = 0.0251$), but the variance explained by the interaction term is negligible (partial $\eta^2 = 0.000608$). The post- vs preinstruction E-CLASS scores for students in courses with the lowest (1) and highest (16) number of weeks spent on open-ended activities according to the linear regression results are plotted in Fig. 4, which confirms that, while additional open-ended activities may have a slightly larger

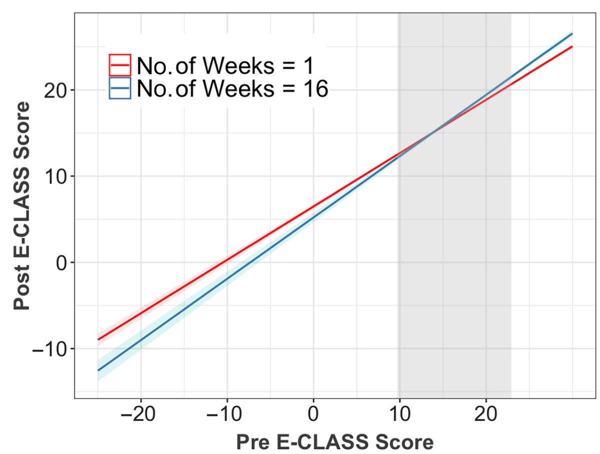

FIG. 4. Expected postinstruction E-CLASS scores (i.e., marginal effects) as a function of preinstruction E-CLASS scores for students in courses with 1 week spent (solid red line) and those with 16 weeks spent (solid blue line) on open-ended activities. The shaded bands around the linear regression lines show one standard deviation intervals. The shaded gray region represents the range of preinstruction scores centered around the global mean with a width of twice the global standard deviation. It is worth noting that the complete E-CLASS dataset is used to fit the model, but we choose to plot the model curves for students only in courses with 1 week and 16 weeks spent on open-ended activities to more easily visualize the outcomes of the fit.

positive effect on postinstruction scores for students with high preinstruction scores, the effect size is too small to be meaningful in an instructional context.

To directly compare E-CLASS scores for varying number of weeks spent on open-ended activities, we removed the interaction term from the model, and reran the analysis without it. We found that the number of weeks spent on open-ended activities is not a significant predictor for the postinstruction score after controlling for the preinstruction score ($\beta_{pre} = 0.02 \pm 0.02,\ p = 0.146$). Additionally, the variance explained by the number of weeks spent on open-ended activities is negligible (partial $\eta^2 = 0.000256$) and remained effectively constant after the interaction term was removed, with any changes falling within the uncertainty of the reported significant figures. We can further confirm this result by plotting the expected postinstruction E-CLASS scores from marginal effects, as shown in Fig. 4, and specifically looking in the range of preinstruction scores centered around the global mean of 16.3 with a width of the global standard deviation of 6.6. Across this region of preinstruction scores, which most student responses (69%) fall within, the two fitted lines cross or are very close together, indicating that for most of these students, the number of weeks spent on open-ended activities is not correlated with pre- to post-instruction gain in E-CLASS score. This finding is consistent with our previous research [21], confirming the absence of a statistically significant correlation between the number of weeks spent on open-ended activities and pre- to post-instruction gain in the students' overall views of experimental physics.

Since we see no evidence of any substantial effect of additional weeks spent on open-ended activities on the postinstruction scores, we turn to the item-by-item analysis using an ordinal logistic regression model to reveal how much adding 1 week of open-ended activities adjusts the likelihood of students' agreement for each individual E-CLASS item after instruction, regardless of how many weeks were already spent on open-ended activities, which again uses the preinstruction score as a covariate and includes an interaction term between preinstruction score and the number of weeks spent on open-ended activities.

After fitting this model to the subset of the E-CLASS matched dataset (excluding courses with $OpenWeeks = 0$), we found that only three out of the 30 E-CLASS items had a significant interaction between the preinstruction score and the number of weeks spent on open-ended activities after the Holm-Bonferroni correction, namely, item 4 ($\tilde{p} = 0.00215$), item 7 ($\tilde{p} = 0.0209$), and item 16 ($\tilde{p} = 0.00387$). We subsequently reran the analysis without the interaction terms to calculate the odds ratios and remind the readers that caution must be taken when interpreting these results, as these two variables impact each other for these three items, leading to the final postinstruction score.

The Holm-Bonferroni corrected p values, odds ratios, and 95% confidence intervals from the ordinal logistic regression for all the individual E-CLASS items are presented in the Appendix B and plotted in Fig. 5. As can be seen in Fig. 5, 6 out of the 30 E-CLASS items demonstrated statistically significant odds ratios after the Holm-Bonferroni correction, and all six of these items have their associated 95% confidence intervals strictly greater than 1. These six items constitute only a small subset of the

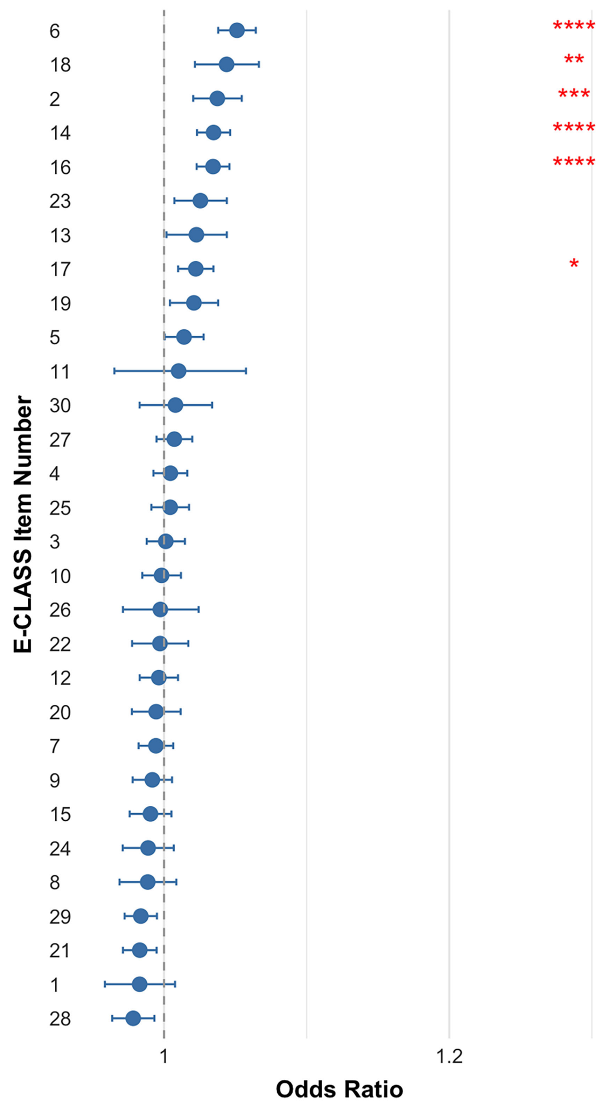

FIG. 5. Odds ratios and their corresponding 95% confidence intervals for students in courses with a varying amount of weeks of open-ended activities. Odds ratios represent the multiplicative change in the odds of moving up a response category. An odds ratio greater than 1 (indicated by the vertical dashed line) signifies a higher likelihood of students in courses with certain number of weeks spent on open-ended activities having a more expertlike response in the post-test for the respective E-CLASS item compared to those in courses with one additional week spent on open-ended activities. Additionally, statistical significance levels are denoted next to each item. Significance at a Holm-Bonferroni corrected p value of less than 0.0001 is represented by ****, p less than 0.001 is ***, p less than 0.01 is **, p less than 0.05 is *, and p values greater than 0.05 are not symbolized.

24 items that exhibit positive impact when comparing open-ended and guided-only activities. The values for these six items with odds ratios greater than one are relatively small, with the largest odds ratio observed being approximately 1.05, which is consistent with the lack of improvement in the overall E-CLASS scores.

Out of these six items, the top five with the largest statistically significant odds ratios for the number of weeks spent on open-ended activities are:

- 6: Scientific journal articles are helpful for answering my own questions and designing experiments.
- 18: Communicating scientific results to peers is a valuable part of doing physics experiments.
- 2: If I wanted to, I think I could be good at doing research.
- 14: When doing an experiment, I usually think up my own questions to investigate.
- 16: The primary purpose of doing physics experiments is to confirm previously known results.

Among these lab courses, those that dedicated a large number of weeks, such as 16, to open-ended activities were likely structured around research projects or similar activities, providing students with greater exposure to authentic scientific practices. For instance, both item 6 and item 18 are associated with scientific communication, indicating that for each additional week of open-ended activities, the odds of students developing more expertlike views about scientific communication increased by about 4%. This means that students in a course with 16 weeks of open-ended activities were roughly twice as likely to develop more expertlike views about scientific communication after instruction than those in a course with 1 week of open-ended activities. Moreover, certain aspects of student confidence in their ability to conduct scientific research, as well as student agency in decision making in their lab courses, also show up in item 2 and item 14, respectively. The improvement on these two aspects specifically are consistent with the findings in previous research on open-ended activities [26,28]. Finally, while item 16 demonstrated a large statistically significant odds ratio for the number of weeks spent on open-ended activities, it is worth reiterating that this item also exhibits a statistically significant interaction between preinstruction score and the number of weeks spent on open-ended activities, so the effect associated with this particular item depends on students' preinstruction scores. The improvement on all of these aspects of experimental physics could be the result of the extended project-based lab courses, and further investigation is needed to confirm the factors contributing to these outcomes.

Overall, the results of the linear regression indicate that the amount of open-ended activities, as measured by the number of weeks, does not significantly impact students' views of experimental physics in general as long as that number is not zero. Moreover, the results of the ordinal logistic regression indicate that additional weeks of open-ended activities could potentially improve the postinstruction score of a small subset of E-CLASS items, regardless of how many weeks already spent on open-ended activities. This finding is consistent with our results on the overall E-CLASS scores, confirming the lack of substantial effect of additional weeks spent on open-ended activities on students' views of experimental physics.

V. CONCLUSIONS

We investigated the correlation between the inclusion of open-ended activities in lab courses and improvement in students' views of experimental physics, as measured by the difference between their overall E-CLASS scores and individual E-CLASS item scores before and after instruction. With an expansive E-CLASS matched dataset, we first treated the courses dichotomously as either having open-ended activities or having only guided activities. Based on the results of the descriptive statistics and ANCOVA on the overall E-CLASS scores, we concluded that lab courses with open-ended activities are associated with more expertlike views of experimental physics as compared to courses that include only traditional guided activities, and that open-ended activities have a larger positive effect on students with less expertlike views before instruction. Through an item-by-item analysis using ordinal logistic regression, we also found that students in courses with open-ended activities are more likely to develop expertlike views surrounding many aspects of experimental physics compared to those in courses with only traditional guided activities after instruction, which aligns with the previous finding on the overall E-CLASS scores.

Subsequently, we constrained our analysis to a subset of the full dataset where we included only the courses with at least 1 week of open-ended activities. Based on the results of the linear regression, we conclude that the number of weeks of open-ended activities is not a predictor of pre- to post-instruction gain in E-CLASS score, a distinct exception to the common trend observed in education research that more instruction results in better student performance [63,64]. This result was reinforced by an item-by-item analysis using ordinal logistic regression, which showed that additional weeks of open-ended activities could potentially improve students' views about a few aspects of experimental physics, but not all aspects. One implicit assumption made here with the use of ordinal logistic regression is that the week-to-week variation is uniform from the lowest (1) to the highest (16) number of weeks spent on open-ended activities, and the odds ratio represents the mean effect of adding 1 week of open-ended activities, regardless of how many weeks already spent on open-ended activities.

The findings of this work call for more research to be conducted in the future to further examine the impact of open-ended activities on students' views of experimental physics. To start with, additional data need to be collected to better characterize the amount and type of open-ended activities included in a course. A source of quantitative data that can be helpful in extracting this information is a growing, international dataset of instructor responses to a new survey that aims to capture the structure, goals, and activities in lab courses around the world, and ultimately create a taxonomy of lab courses that can be used by both instructors and education researchers [65]. In particular, the survey contains an "activity" section, where instructors are asked to report the frequency of data-analysis-and-visualization, communication, student-decision-making, materials-and-resources, as well as modeling-based activities in their courses [65]. A source of more qualitative data is classroom observation transcripts and videos, which include a wealth of information regarding student-instructor interaction, student teamwork, etc., and can be thematically coded to categorize the nature of open-ended activities and describe how students engage with them.

Furthermore, several factors were not controlled for in this study, but could act as mediators, confounders, or colliders of the relationship between open-ended activities and students' views of experimental physics [66]. For instance, this work did not take into account the lab courses students had taken before the course under this study, nor the types of activities included in the guided portion of the course, both of which could affect students' views of experimental physics. One additional factor that we did not control for is the instructors' attitude and motivation. Instructors who include at least 1 week of open-ended activities may value improving students' views of experimental physics differently than those who do not. As such, even for courses with a low number of weeks spent on open-ended activities, the guided portion may still have more open-endedlike pedagogical features, or include activities that still engage students in authentic research practices. Another factor to explore is the class size and student-to-instructor ratio. Past research has shown that positive shifts in students' views tend to occur in small classes [67]. It is likely that even with a low number of weeks spent on open-ended activities, courses that encourage student-instructor interaction and student teamwork, either from having small class sizes or low student-to-instructor ratios, could help students develop more expertlike views of experimental physics.

Additionally, as discussed earlier, this study is exploratory and descriptive in nature, and we have employed only correlational analysis methods and predictive models to answer our research questions [66]. We do not have evidence for any causation. Future studies are needed to replicate the results of this study, and establish the causal relationship between the amount of open-ended activities and improvement in students' views of experimental physics. A purely quantitative study could involve a quasiexperimental design where students in the same course were provided with two different types of instruction, where the control condition offers limited open-ended activities, and the treatment condition offers additional open-ended activities. However, this design might be difficult to implement in a classroom setting. Therefore, the application of qualitative methods, such as interviews with students and instructors on their perspectives on the inclusion of open-ended activities, and classroom observational protocol that encapsulates the instructors' implementation of, and the students' engagement in, the open-ended activities could be helpful or even necessary in supplementing a purely quantitative study.

Finally, while we have focused on how open-ended activities may improve students' views of experimental physics, there are many other potential benefits of open-ended activities, such as improving students' experimental skills, that can contribute significantly to student learning of experimental physics. Future work is needed to explore these other facets.

Nevertheless, this study contributes to the broader understanding of open-ended activities' positive impact in lab courses and demonstrates that a large amount of open-ended activities may not be necessary to improve students' views of experimental physics. This has the potential to significantly lower the barrier for instructors to include open-ended activities in their lab courses in an effort to improve students' views of experimental physics.

ACKNOWLEDGMENTS

We would like to thank Bethany Wilcox for her help and feedback. This research was supported by NSF DUE-2142356 and PHY 2317149. Publication of this article was partially funded by the University of Colorado Boulder Libraries Open Access Fund.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX A: HISTOGRAMS OF FRACTION OF WEEKS SPENT ON OPEN-ENDED ACTIVITIES

Figure 6 presents the distributions of fraction of weeks spent on open-ended activities by course and students enrolled. These histograms are for informational purposes only, as we do not use them in our analysis. Our choice to parameterize the amount of open-ended instruction was grounded in theoretical considerations. For example, a fully open-ended semester-long course and a fully open-ended quarter-long course may have the same proportion of open-ended weeks, but students may experience 15 weeks of open-ended instruction very differently from just eight. To better capture this difference, we focus on the concept of time-on-task to quantify open-ended instruction, rather than including the number of weeks of prescriptive teaching in the parameterization.

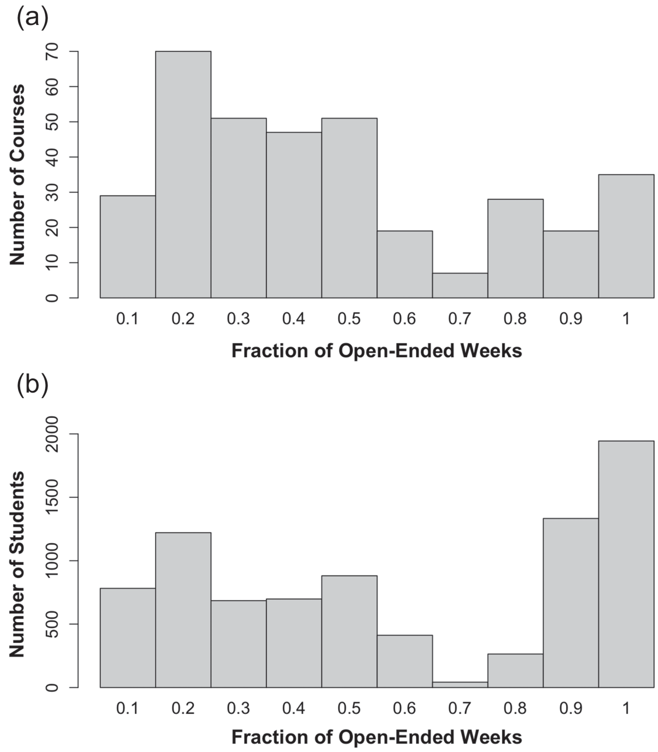

FIG. 6. Histograms of the fraction of weeks spent on open-ended activities for (a) the 356 lab courses that include at least 1 week of open-ended activities and (b) the 8262 students that are enrolled in these courses.

APPENDIX B: TABLE OF RESULTS FROM LOGISTIC REGRESSION MODELS

Table III presents individual E-CLASS items along with their item number referenced in Figs. 3 and 5. The Holm-Bonferroni corrected p values ($\tilde{p}$), odds ratios (OR), and 95% confidence intervals (CI) from the ordinal logistic regression models of each item for the two research questions, RQ1 and RQ2, are also provided.

TABLE III. Individual E-CLASS items, odds ratios (OR), 95% confidence intervals (CI), and p values (p) of the ordinal logistic regression models for the two research questions, RQ1 and RQ2.

		RQ1: $OpenEnded$			RQ2: $OpenWeeks$		
No.	E-CLASS item	$\tilde{p}$	OR	95% CI	$\tilde{p}$	OR	95% CI
1	When doing an experiment, I try to understand how the experimental setup works.	3.83×10^{-16}	1.548	[1.400, 1.714]	1.000	0.985	[0.963, 1.007]
2	If I wanted to, I think I could be good at doing research.	1.42×10^{-9}	1.245	[1.165, 1.331]	4.52×10^{-4}	1.035	[1.019, 1.051]
3	When doing a physics experiment, I do not think much about sources of systematic error.	8.76×10^{-18}	1.288	[1.218, 1.361]	1.000	1.001	[0.989, 1.013]
4	If I am communicating results from an experiment, my main goal is to have the correct sections and formatting.	2.34×10^{-18}	1.266	[1.203, 1.332]	1.000	1.004	[0.993, 1.015]
5	Calculating uncertainties usually helps me understand my results better.	2.75×10^{-7}	1.172	[1.108, 1.239]	0.789	1.013	[1.000, 1.026]
6	Scientific journal articles are helpful for answering my own questions and designing experiments.	6.91×10^{-17}	1.264	[1.199, 1.332]	8.72×10^{-13}	1.048	[1.035, 1.060]
7	I do not enjoy doing physics experiments.	7.80×10^{-14}	1.225	[1.165, 1.289]	1.000	0.995	[0.984, 1.006]

(Table continued)

TABLE III. *(Continued)*

		RQ1: *OpenEnded*			RQ2: *OpenWeeks*		
No.	E-CLASS item	$\tilde{p}$	OR	95% CI	$\tilde{p}$	OR	95% CI
8	When doing an experiment, I try to understand the relevant equations.	2.47×10^{-12}	1.365	[1.257, 1.484]	1.000	0.990	[0.972, 1.008]
9	When I approach a new piece of lab equipment, I feel confident I can learn how to use it well enough for my purposes.	1.000	1.018	[0.961, 1.079]	1.000	0.993	[0.980, 1.005]
10	Whenever I use a new measurement tool, I try to understand its performance limitations.	3.99×10^{-21}	1.323	[1.251, 1.400]	1.000	0.998	[0.986, 1.011]
11	Computers are helpful for plotting and analyzing data.	6.02×10^{-4}	1.449	[1.209, 1.746]	1.000	1.009	[0.969, 1.054]
12	I do not need to understand how the measurement tools and sensors work in order to carry out an experiment	1.000	1.002	[0.947, 1.061]	1.000	0.997	[0.985, 1.009]
13	If I try hard enough, I can succeed at doing physics experiments.	1.000	1.008	[0.925, 1.099]	0.670	1.021	[1.002, 1.041]
14	When doing an experiment, I usually think up my own questions to investigate.	3.05×10^{-82}	1.627	[1.549, 1.709]	1.35×10^{-7}	1.032	[1.021, 1.043]
15	Designing and building things is an important part of doing physics experiments.	5.37×10^{-23}	1.376	[1.295, 1.464]	1.000	0.991	[0.978, 1.005]
16	The primary purpose of doing physics experiments is to confirm previously known results.	2.96×10^{-30}	1.336	[1.273, 1.402]	1.35×10^{-7}	1.032	[1.021, 1.043]
17	When I encounter difficulties in the lab, my first step is to ask an expert, like the instructor.	2.95×10^{-37}	1.428	[1.354, 1.507]	0.0106	1.020	[1.009, 1.032]
18	Communicating scientific results to peers is a valuable part of doing physics experiments.	1.62×10^{-13}	1.401	[1.287, 1.526]	3.31×10^{-3}	1.041	[1.020, 1.063]
19	Working in a group is an important part of doing physics experiments.	5.32×10^{-10}	1.258	[1.175, 1.347]	0.302	1.019	[1.004, 1.035]
20	I enjoy building things and working with my hands.	1.000	1.050	[0.980, 1.126]	1.000	0.995	[0.980, 1.011]
21	I am usually able to complete an experiment without understanding the equations and physics ideas that describe the system I am investigating.	3.45×10^{-13}	1.214	[1.155, 1.276]	0.108	0.985	[0.974, 0.995]
22	If I am communicating results from an experiment, my main goal is to make conclusions based on my data using scientific reasoning.	0.0121	1.142	[1.051, 1.241]	1.000	0.997	[0.980, 1.016]
23	When I am doing an experiment, I try to make predictions to see if my results are reasonable.	3.79×10^{-23}	1.460	[1.359, 1.571]	0.138	1.024	[1.007, 1.041]
24	Nearly all students are capable of doing a physics experiment if they work at it.	1.000	1.027	[0.953, 1.107]	1.000	0.990	[0.974, 1.006]
25	A common approach for fixing an experiment is to randomly change things until the problem goes away.	1.000	1.018	[0.963, 1.076]	1.000	1.004	[0.992, 1.016]
26	It is helpful to understand the assumptions that go into making predictions.	3.34×10^{-17}	1.608	[1.448, 1.790]	1.000	0.998	[0.974, 1.022]
27	When doing an experiment, I just follow the instructions without thinking about their purpose.	6.65×10^{-30}	1.363	[1.294, 1.436]	1.000	1.007	[0.995, 1.018]
28	I do not expect doing an experiment to help my understanding of physics.	2.67×10^{-12}	1.259	[1.184, 1.339]	0.101	0.980	[0.967, 0.994]

(Table continued)

TABLE III. *(Continued)*

		RQ1: *OpenEnded*			RQ2: *OpenWeeks*		
No.	E-CLASS item	$\tilde{p}$	OR	95% CI	$\tilde{p}$	OR	95% CI
29	If I do not have clear directions for analyzing data, I am not sure how to choose an appropriate analysis method.	3.40×10^{-30}	1.336	[1.273, 1.402]	0.108	0.985	[0.975, 0.995]
30	Physics experiments contribute to the growth of scientific knowledge.	3.86×10^{-9}	1.385	[1.252, 1.536]	1.000	1.007	[0.985, 1.031]

[1] R. Trumper, The physics laboratory—A historical overview and future perspectives, Sci. Educ. **12**, 645 (2003).

[2] S. B. McKagan, D. A. Craig, M. Jackson, and T. Hodapp, *A Guide to Effective Practices for Physics Programs (EP3)* (American Physical Society, College Park, MD, 2021), https://ep3guide.org/.

[3] N. G. Holmes and C. E. Wieman, Introductory physics labs: We can do better, Phys. Today **71**, No. 1, 38 (2018).

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

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

[6] B. M. Zwickl, N. Finkelstein, and H. J. Lewandowski, The process of transforming an advanced lab course: Goals, curriculum, and assessments, Am. J. Phys. **81**, 63 (2013).

[7] D. Doucette, B. D'Urso, and C. Singh, Lessons from transforming second-year honors physics lab, Am. J. Phys. **88**, 838 (2020).

[8] J. R. Hoehn, M. F. Fox, A. Werth, V. Borish, and H. Lewandowski, Remote advanced lab course: A case study analysis of open-ended projects, Phys. Rev. Phys. Educ. Res. **17**, 020111 (2021).

[9] E. Etkina and A. Van Heuvelen, Investigative science learning environment—A science process approach to learning physics, in *Proceedings of the 2001 Physics Education Research Conference* (American Association of Physics Teachers, College Park, MD, 2001).

[10] M. Wells, D. Hestenes, and G. Swackhamer, A modeling method, Am. J. Phys. **63**, 606 (1995).

[11] R. J. Beichner, J. M. Saul, R. J. Allain, D. L. Deardorff, and D. S. Abbott, Introduction to scale-up: Student-centered activities for large enrollment university physics, in *Proceedings of the Annual Meeting of the American Society for Engineering Education* (American Society for Engineering Education, Washington, DC, 2000).

[12] R. E. Scherr and N. Holmes, Quantifying uncertainty and distinguishing data sets in introductory physics, CourseSource **10** (2023).

[13] A. Werth, K. Oliver, C. G. West, and H. Lewandowski, Assessing student engagement with teamwork in an online, large-enrollment course-based undergraduate research experience in physics, Phys. Rev. Phys. Educ. Res. **18**, 020128 (2022).

[14] A. Werth, C. G. West, and H. Lewandowski, Impacts on student learning, confidence, and affect in a remote, large-enrollment, course-based undergraduate research experience in physics, Phys. Rev. Phys. Educ. Res. **18**, 010129 (2022).

[15] R. Merritt and H. Lewandowski, Physics instructor views on course-based undergraduate research experiences (CUREs), *presented at PER Conf. 2024 Boston, MA*, 10.1119/perc.2024.pr.Merritt.

[16] M. Kretchmer, R. Merritt, and H. Lewandowski, Exploring student beliefs of traditional physics laboratory coursework in relation to authentic research, *presented at PER Conf. 2024 Boston, MA*, 10.1119/perc.2024.pr.Kretchmer.

[17] B. R. Wilcox, Students' views about the nature of experimental physics, Phys. Rev. Phys. Educ. Res. **13**, 020110 (2017).

[18] B. R. Wilcox and H. Lewandowski, A summary of research-based assessment of students' beliefs about the nature of experimental physics, Am. J. Phys. **86**, 212 (2018).

[19] B. M. Zwickl, T. Hirokawa, N. Finkelstein, and H. J. Lewandowski, Epistemology and expectations survey about experimental physics: Development and initial results, Phys. Rev. ST Phys. Educ. Res. **10**, 010120 (2014).

[20] B. R. Wilcox and H. J. Lewandowski, Students' epistemologies about experimental physics: Validating the Colorado learning attitudes about science survey for experimental physics, Phys. Rev. Phys. Educ. Res. **12**, 010123 (2016).

[21] Q. Liu and H. Lewandowski, Correlation of open-ended activities in laboratory courses with students' views of experimental physics, *presented at PER Conf. 2024 Boston, MA*, 10.1119/perc.2024.pr.Liu.

[22] A. R. Carter, One hundred years later, introductory labs are poised for change, Phys. Teach. **59**, 97 (2021).

[23] National Research Council, Center for Science, Mathematics, and Engineering Education, and Committee on Development of an Addendum to the National Science Education Standards on Scientific Inquiry, *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning* (Center for Science, Mathematics, and Engineering Education/National Academy Press, Washington, DC, 2000).

[24] J. Kozminski, H. Lewandowski, N. Beverly, S. Lindaas, D. Deardorff, A. Reagan, R. Dietz, R. Tagg, M. EblenZayas, J. Williams *et al.*, AAPT recommendations for the undergraduate physics laboratory curriculum, Phys. Teach. **53**, 253 (2014).

[25] A. Bandura, Human agency in social cognitive theory, Am. Psychol. **44**, 1175 (1989).

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

[27] J. A. Schmidt, J. M. Rosenberg, and P. N. Beymer, A person-in-context approach to student engagement in science: Examining learning activities and choice, J. Res. Sci. Teach. **55**, 19 (2018).

[28] E. Jeffery, K. Nomme, T. Deane, C. Pollock, and G. Birol, Investigating the role of an inquiry-based biology lab course on student attitudes and views toward science, CBE Life Sci. Educ. **15**, ar61 (2016).

[29] D. R. Dounas-Frazer, J. T. Stanley, and H. Lewandowski, Student ownership of projects in an upper-division optics laboratory course: A multiple case study of successful experiences, Phys. Rev. Phys. Educ. Res. **13**, 020136 (2017).

[30] E. M. Smith, M. M. Stein, C. Walsh, and N. Holmes, Direct measurement of the impact of teaching experimentation in physics labs, Phys. Rev. X **10**, 011029 (2020).

[31] H. A. Schweingruber, M. L. Hilton, and S. R. Singer, *America's Lab Report: Investigations in High School science* (National Academies Press, Washington, DC, 2006).

[32] K. Henige, Undergraduate student attitudes and perceptions toward low-and high-level inquiry exercise physiology teaching laboratory experiences, Adv. Physiol. Educ. **35**, 197 (2011).

[33] D. C. Owens, T. D. Sadler, A. T. Barlow, and C. Smith-Walters, Student motivation from and resistance to active learning rooted in essential science practices, Res. Sci. Educ. **50**, 253 (2020).

[34] Z. Y. Kalender, E. Stump, K. Hubenig, and N. Holmes, Restructuring physics labs to cultivate sense of student agency, Phys. Rev. Phys. Educ. Res. **17**, 020128 (2021).

[35] B. M. Zwickl, T. Hirokawa, N. Finkelstein, and H. Lewandowski, Development and results from a survey on students views of experiments in lab classes and research, *presented at PER Conf. 2014, Portland, OR*, 10.1119/perc.2013.pr.083.

[36] J. M. Aiken and H. Lewandowski, Data sharing model for physics education research using the 70 000 response Colorado learning attitudes about science survey for experimental physics dataset, Phys. Rev. Phys. Educ. Res. **17**, 020144 (2021).

[37] N. Holmes and H. Lewandowski, Investigating the landscape of physics laboratory instruction across North America, Phys. Rev. Phys. Educ. Res. **16**, 020162 (2020).

[38] D. Doucette, R. Clark, and C. Singh, Students' attitudes toward experimental physics in a conceptual inquiry-based introductory physics lab, Can. J. Phys. **100**, 292 (2022).

[39] N. Sulaiman, A. Werth, and H. Lewandowski, Students' views about experimental physics in a large-enrollment introductory lab focused on experimental scientific practices, Phys. Rev. Phys. Educ. Res. **19**, 010116 (2023).

[40] A. Werth, C. G. West, N. Sulaiman, and H. Lewandowski, Enhancing students' views of experimental physics through a course-based undergraduate research experience, Phys. Rev. Phys. Educ. Res. **19**, 020151 (2023).

[41] M. F. Fox, J. R. Hoehn, A. Werth, and H. Lewandowski, Lab instruction during the COVID-19 pandemic: Effects on student views about experimental physics in comparison with previous years, Phys. Rev. Phys. Educ. Res. **17**, 010148 (2021).

[42] B. R. Wilcox and H. Lewandowski, Correlating students' beliefs about experimental physics with lab course success, *presented at PER Conf. 2015 College Park, MD*, 10.1119/perc.2015.pr.087.

[43] B. R. Wilcox and H. Lewandowski, Impact of instructional approach on students' epistemologies about experimental physics, *presented at PER Conf. 2016 Sacramento, CA*, 10.1119/perc.2016.pr.092.

[44] B. R. Wilcox and H. Lewandowski, Research-based assessment of students' beliefs about experimental physics: When is gender a factor?, Phys. Rev. Phys. Educ. Res. **12**, 020130 (2016).

[45] 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).

[46] B. R. Wilcox and H. Lewandowski, Improvement or selection? A longitudinal analysis of students' views about experimental physics in their lab courses, Phys. Rev. Phys. Educ. Res. **13**, 023101 (2017).

[47] C. Walsh, H. Lewandowski, and N. Holmes, Skills-focused lab instruction improves critical thinking skills and experimentation views for all students, Phys. Rev. Phys. Educ. Res. **18**, 010128 (2022).

[48] M. Lovelace and P. Brickman, Best practices for measuring students' attitudes toward learning science, CBE Life Sci. Educ. **12**, 606 (2013).

[49] I. A. Halloun, Student views about science, in *Proceedings of the International Conference on Undergraduate Physics Education, Educational Research Center, Lebanese University* (American Institute of Physics Press, College Park, MD, 2001).

[50] B. R. Wilcox, B. M. Zwickl, R. D. Hobbs, J. M. Aiken, N. M. Welch, and H. Lewandowski, Alternative model for administration and analysis of research-based assessments, Phys. Rev. Phys. Educ. Res. **12**, 010139 (2016).

[51] F. Wilcoxon, Individual comparisons by ranking methods, in *Breakthroughs in Statistics: Methodology and Distribution* (Springer, New York, 1992), pp. 196–202.

[52] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Routledge, New York, 2013).

[53] B. Thompson, What future quantitative social science research could look like: Confidence intervals for effect sizes, Educ. Res. **31**, 25 (2002).

[54] B. Thompson, Effect sizes, confidence intervals, and confidence intervals for effect sizes, Psychologie **44**, 423 (2007).

[55] A. Wildt, *Analysis of Covariance* (Sage, Thousand Oaks, CA, 1978).

[56] A. Gelman, *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge University Press, Cambridge, England, 2007).

[57] S. Holm, A simple sequentially rejective multiple test procedure, Scand. J. Stat. Theory Appl. **6**, 65 (1979).

[58] M. Aickin and H. Gensler, Adjusting for multiple testing when reporting research results: The Bonferroni vs Holm methods, Am. J. Public Health **86**, 726 (1996).

[59] G. Van Rossum, F. L. Drake *et al.*, *Python Reference Manual* (Centrum voor Wiskunde en Informatica, Amsterdam, 1995), Vol. 111.

[60] R Core Team, *R: A Language and Environment for Statistical Computing* (R foundation for statistical Computing, Vienna, Austria, 2020).

[61] D. L. Hahs-Vaughn and R. G. Lomax, *An Introduction to Statistical Concepts* (Routledge, New York, 2020).

[62] G. A. Miller and J. P. Chapman, Misunderstanding analysis of covariance, J. Abnorm. Psychol. **110**, 40 (2001).

[63] D. Clark and M. C. Linn, Designing for knowledge integration: The impact of instructional time, J. Learn. Sci. **12**, 451 (2003).

[64] S. C. Andersen, M. K. Humlum, and A. B. Nandrup, Increasing instruction time in school does increase learning, Proc. Natl. Acad. Sci. U.S.A. **113**, 7481 (2016).

[65] G. Geschwind, M. Alemani, M. F. Fox, P. Logman, E. Tufino, and H. Lewandowski, Development of a global landscape of undergraduate physics laboratory courses, Phys. Rev. Phys. Educ. Res. **20**, 020117 (2024).

[66] V. Adlakha and E. Kuo, Critical issues in statistical causal inference for observational physics education research, Phys. Rev. Phys. Educ. Res. **19**, 020160 (2023).

[67] A. Madsen, S. B. McKagan, and E. C. Sayre, How physics instruction impacts students' beliefs about learning physics: A meta-analysis of 24 studies, Phys. Rev. ST Phys. Educ. Res. **11**, 010115 (2015).