Comparison of labatorials and traditional labs: The impacts of instructional scaffolding on the student experience and conceptual understanding

Franco La Braca[*] and Calvin S. Kalman[†]

Department of Physics, Concordia University, Montreal, Quebec H4B 1R6, Canada

Traditional, template physics labs are often associated with student dissatisfaction and superficial applications, and are known to leave students with fragmented knowledge. An alternative known as labatorials, a conceptually driven approach to labs, has been proposed. In a number of studies, labatorials have been shown to work well. However, what has been missing is a study comparing labatorials to traditional labs. In this study, labatorials are compared with traditional labs in terms of students' learning experience and the quality of their conceptual learning. Additionally, we identify the scaffolding mechanisms that impact these elements. In the context of Concordia University's introductory experimental mechanics course, we collect data spanning semistructured student and teaching assistant (TA) interviews, class observations, TA surveys, post-test and final exam scores and responses, and student writing products. Upon analysis and triangulation, we find that due to the scaffolding present in labatorials, students typically exhibit a high degree of collaboration and engagement with the material in a low-pressure environment, which allows students to focus on the learning. This is attributed to three primary forms of scaffolding inherent to the design of labatorials: peer scaffolding, instructor scaffolding, and scaffolding by the activity worksheet. In contrast, students in traditional labs have a tendency to rely on step-by-step instructions and focus on avoiding errors, which may inhibit their conceptual learning. These conclusions are supported by the students' differing performance and understanding exhibited in different types of questions; traditional lab students tend to perform better on questions involving standardized processes or simple, memorization-based calculations, while labatorial students tend to perform better on conceptual questions.

I. INTRODUCTION

A. The physics laboratory

The physics lab has long been a distinctive part of physics education [1–5]. Physics labs play a central role in teaching and learning physics at the high school and undergraduate levels due to the opportunity they can give students to participate in authentic inquiry activities and investigations. However, for a long time, education research had not been able to show that physics labs were meeting this potential, bringing the long-standing assumption of their usefulness into question [6]. Even in the 21st century, where science education researchers have made notable strides in understanding what constitutes a well-designed laboratory, there exist large discrepancies between the recommendations of researchers and what is being implemented by practitioners [7]. Because of diverse constraints, many experiments are still being framed as a recipelike sequence of steps, resulting in students thinking neither conceptually about the phenomenon under investigation nor critically about the experimental procedure.

As a result, students often do not perceive the purpose of physics labs, especially at the introductory level. It is our experience that they believe labs to be boring and tedious exercises that do nothing but culminate in a cumbersome report. Students just focus on following the recipelike instructions, collect the necessary data, and attempt to piece together their understanding of the experiment once they begin writing the lab report at home [2,8]. Students are left with fragmented knowledge, leaving the course none the wiser about the relevant concepts [6,9–12]. Reflective of this, in a study of 3 institutions, taught by 7 instructors, and taken by nearly 3000 students with a high degree of precision, there was no statistically measurable benefit on student course performance from enrolling in the associated lab course [12]. Traditional labs also typically do not foster creativity in methodology or experimental design, which limits the growth of students' experimental skills [3]. Moreover, students are typically dissatisfied with the traditional physics lab experience, often expressing that it is boring or irrelevant to real life [13–15].

---
[*]franco.labraca@concordia.ca
[†]calvin.kalman@concordia.ca

Labatorials (combination of "lab" and "tutorial") are an alternative approach to physics labs that aims to alleviate these common concerns about traditional labs. Labatorials aim to foster students' conceptual understanding as they perform experiments. Labatorials were developed at the University of Calgary [16] inspired by the introductory physics tutorial system entitled "Tutorials in Introductory Physics" at the University of Washington [17]. Labatorials use misconception-targeting worksheets similar in structure to those in tutorials, except with additional emphasis on experiments; labatorial worksheets are designed so as to encourage an ongoing interaction between the students and lab instructor as they proceed, allowing them to receive immediate feedback. This scaffolded approach to learning and experimentation aims to help students develop a deep understanding of physics concepts beyond just being able to perform calculations or execute procedures.

Similarly to tutorials, students in labatorials work collaboratively in teams using a worksheet that gradually builds up their understanding of the material while making use of peer instruction [18,19] and instructor guidance at key locations across the worksheet. In addition to being able to learn in a supportive environment, the worksheet also actively aims to connect the material to students' lives, giving purpose to the concepts studied. Because of the nature of the worksheet and the main goal of a labatorial, a follow-up lab report is not mandatory.

Although distinct in implementation, the role of scaffolding in labatorials is similar to that in the investigative science learning environment (ISLE) system [20] in that both are inquiry-based and are designed with appropriate scaffolding to enable students to construct knowledge through the investigation of physical phenomena. However, while ISLE aims to address the reported issues with traditional labs by letting students participate in genuine inquiry cycles, fostering students' interpretive knowing skills, labatorials emphasize developing conceptual understanding of the phenomena under investigation.

B. Theoretical framework

The underlying theoretical framework of labatorials is that of guided inquiry. Guided inquiry is a pedagogical approach based on the core ideas of social constructivism and inquiry-based learning.

Social constructivism addresses the "guided" element of the guided inquiry framework. In particular, social constructivist (or sociocultural) theories of learning were proposed by Bruner [21–23] and are based on the work of Vygotsky [24]. The pivotal idea of all forms of constructivism as a cognitive theory is that learners actively construct new knowledge from experiences. In social constructivism, emphasis is placed on the role that the relationships and interactions between the learner and more expertlike individuals play in the learner's knowledge construction process, in particular on the more

knowledgeable individual's role in scaffolding the learning experience.

Instructional scaffolding, based on the ideas of Vygotsky and coined by Bruner, is the way that complex knowledge structures are constructed with the support of appropriate guidance and resources put in place by those with more expertlike knowledge. By installing appropriate temporary structural supports, more permanent, stronger structures can be built up into the eventual final product, at which point the scaffolding is no longer required. Such scaffolding is to be distinguished from simple telling or instructing, where the expert simply tries to transmit their knowledge to the student; the student performs the mental task independently, but with appropriate guidance as provided by the expert.

We conjecture that scaffolding also plays a vital role in resolving the cognitive dissonance often experienced by students when trying to learn a new concept. Specifically, cognitive dissonance occurs when students encounter concepts that differ from their perceptions of the world [25]. This is particularly pertinent to Newtonian mechanics since students have been experiencing the world since a young age, from which they have been developing their own intuition of how and why objects move. These strong, and often incorrect, prior notions about physics can cause students to warp their perception of new information so as to reduce the associated cognitive dissonance, thereby interfering with conventional instruction. However, with an approach purposefully designed to target well-known misconceptions (e.g., tutorials and labatorials) that guides students' processing of new information so as to confront and resolve inconsistencies, cognitive dissonance can be overcome, allowing them to correctly learn a new concept.

The idea of inquiry-based learning is intrinsically linked to the core ideas of constructivism such as scaffolding and, if utilized appropriately, can also assist in the resolution of misconceptions. In this approach, it is posited that if students are to construct their own knowledge, then they should be experientially learning; it is by actively participating in personal or authentic experiences that students will be able to derive the most meaning from their learning [26,27]. This meaning-making process thus elicits the cognitive engagement of students with the course material through investigation and collaboration.

When speaking of cognitive engagement in this work, we are referring to the definition of Rotgans and Schmidt [28], who define the concept as "the extent to which students are willing and able to take on the learning task at hand." Therefore, high cognitive engagement occurs when a student is capable of contributing autonomously to a task and is therefore willing to put in the effort and time required to complete the task. In the context of conceptual understanding in labs, this interpretation can further be specified to refer to engagement with the concepts of the lab; if a student is cognitively engaged, they will be

constantly thinking and putting in the effort resolve any cognitive dissonance encountered as they grapple with the concepts.

To this end, in labatorials students are encouraged to not passively proceed without understanding. In particular, labatorials utilize learning processes that are characteristic of the general approach in any form of inquiry-based learning [29,30]. Namely, students should

(1) Formulate their own questions.
(2) Collect data as evidence to answer the questions.
(3) Develop hypotheses based on the evidence collected.
(4) Make connections between the knowledge obtained while investigating and the proposed hypotheses.
(5) Make an argument to justify the hypotheses.

In contrast to verification labs, an inquiry-based lab helps students partake in an authentic process of scientific investigation through the above steps, which helps them attain a deeper understanding of the course content. What distinguishes guided inquiry from other forms of inquiry [31,32] is the level to which the aforementioned steps are scaffolded for the students. In the guided approach, the instructor provides the research question for the students. However, students are still expected to design an appropriate protocol for investigating the question, collecting data, and presenting their findings.

The labatorial literature suggests not only the existence of such scaffolding, but also the existence of scaffolding mechanisms for facilitating conceptual understanding. The labatorial approach was first described in detail by Ahrensmeier *et al.* [33] and Ahrensmeier [8], who suggested improvements in introductory-level university students' attitudes toward physics as well as their conceptual understanding and problem solving skills. The impact of labatorials in conjunction with reflective writing [34], an activity that allows students to metacognitively examine textual material, on students' epistemological beliefs was also explored by Kalman *et al.* [35], and it was found that with such a combination of interventions, students' epistemological beliefs could become more expertlike. Sobhanzadeh, Kalman, and Thompson [36] showed that the structure of labatorials encouraged students to interact more frequently with their peers and the TA as well as get more deeply engaged in the lab. This had a positive impact on their conceptual understanding and overall experience compared to prior implementations of the course, with improved student satisfaction and lowered student anxiety. Similar results were also found by Kalman, El-Helou, and Lattery [37] when labatorials, which were originally designed for introductory university physics courses, were implemented at the high school level, particularly emphasizing the relative ease of implementation of the approach.

What has been missing is a study comparing labatorials to traditional labs. Such a comparison is the goal of this paper. In the context of social constructivist ideas of scaffolding, we hypothesize that the prior successes of labatorials were due to several forms of scaffolding in the overall design of labatorials. Despite having been created as an alternative to traditional labs, all labatorial-related studies conducted thus far have only considered labatorials in and of themselves, without side-by-side comparison to a traditional laboratory equivalent of the same course in a controlled experimental design. In pursuing the goal of this study, we explicitly identify and characterize the scaffolding mechanisms of labatorials so as to allow us to compare and contrast students' learning experience and conceptual understanding in labatorials and traditional labs in terms of these mechanisms.

II. METHODOLOGY

In order to compare labatorials and traditional labs with regard to the students' learning experience and conceptual understanding, we set out to answer the following two primary research questions and their component subquestions (which we will henceforth refer to using shorthand notation beginning with "RQ"):

(1) How can the learning experience differ between labatorials and traditional labs?
   (a) How do social interactions in the lab impact the learning experience?
   (b) What elements of labs play a role in providing a satisfying learning experience?
  2. In what ways do labatorials and traditional labs promote the development of conceptual understanding?
   (a) What elements of physics labs help students improve their conceptual understanding?
   (b) How do students' learning outcomes compare between the two lab approaches?

When using the term "learning outcome" in this work, we are referring to the definition of Biggs and Collis [38], i.e., a concept or skill that has been mastered (or not) by a student by the end of the course as indicated by their performance. Although the term learning outcome is often used in the field of curricular design to refer to a formal statement specifying the aforementioned concepts or skills to be mastered, we use the stated definition to emphasize the notion of students' understanding being the result (or outcome) of instruction, which is more pertinent to our research question.

For this study, we adopt a mixed methods design with a focus on qualitative analysis. The design involves concurrent qualitative and quantitative data collection, which is integrated at the data interpretation phase of the research using a concurrent triangulation strategy. Differently stated, we first complete all types of data collection (whether qualitative or quantitative), then analyze each type of data separately, and finally corroborate the qualitative and quantitative data sources (if appropriate) in order to derive a meaningful interpretation of the results.

Although most of the research questions are qualitative in nature and therefore well suited to qualitative analysis, RQ2b (research question 2b) will also be explored quantitatively as students' learning outcomes can affect their performance on assessments. Additionally, meaningful comparison of two groups post-instruction requires establishing that the groups are initially equivalent, which can be done statistically. These considerations necessitate the use of both qualitative and quantitative methods in this study. However, it is our experience that traditional assessments do not always successfully capture students' level of conceptual understanding. As such, although we wish to quantitatively explore the course assessments for possible insights into the data and to direct our qualitative analysis as well as lay the foundation for meaningful comparison, our global arguments and conclusions will be qualitatively driven, with the quantitative data playing only a supportive role in the analysis and triangulation.

A. Context of the study

The study takes place in the context of the introductory experimental mechanics course at Concordia University (Montreal, Canada). This is a one-credit, freshmen level lab course (the first in a sequence of introductory labs) that must be taken either concurrently with or after having completed the introductory mechanics lecture course.

Although there is a major pedagogical challenge associated with this lab course in that many students do not take the associated lecture course concurrently, this is actually an advantage for this study. There is less contamination of ideas expressed in the lecture with those investigated in the lab.

As with most universities offering such courses, both the lab and lecture courses are service courses, i.e., they are required of all students wishing to pursue a science major. Furthermore, nonscience students also often take this course either to acquire an extra credit for their degree requirements or to learn something new outside their field.

As such, the student population of the course is typically very heterogeneous in terms of gender, background, and goals. There are six two- to three-hour experiments to be performed in the course. In the traditional version of the course, students work in self-assembled teams of two to four for labs 1 and 6, but work individually for the other labs. Furthermore, there are no pre- or post-lab activities, and students are given instructions to follow for each experiment as well as key points to address in their reports. For writing lab reports, students are given a template filled with an example to follow at the beginning of their lab manual. Reports are short (one to two pages) and written at the end of class, typically handed in directly to the teaching assistant (TA) overseeing the lab section and then returned to the students the following class.

There were only 54 students enrolled in total. The experimental group, i.e., the labatorial lab students, consisted of 30 students across three course sections (sections 40, 42, and 44 containing 11, 9, and 10 students, respectively) and the control group, i.e., the traditional lab students, consisted of 24 students across two course sections (sections 41 and 43 each containing 12 students). This count of students includes only the students who remained in the course from beginning to end, which are the students for whom we have complete quantitative data. A subset of the students in the course participated in two interviews, and so of those students, only those who participated in both were included in the study sample. Students were placed by the registrar's office into one of the five course sections upon registering for the course. Although the students were not given the choice before the course began, all students were made aware of the details of the study and were asked to give consent to participate through a consent form, and arrangements could have been made on a case-by-case basis had anyone opposed. Only those who consented were included in the study sample. Students could then volunteer to participate in the interviews. Although interviewees were solicited

TABLE I. Labatorial and traditional lab interviewee metadata.

Pseudonym	Interviewee characteristic		
	Section	Major	Prior physics experience
Catherine	40	Biology	10 years ago in HS
Quincy	40	Environmental Science	Recently in college
Emma	42	Exercise Science	No physics in HS
Derek	44	Behavioral Neuroscience	Recently in college
Jessica	44	Exercise Science	10 years ago in HS
Stacy	44	Biochemistry	10 years ago in university
Adrian	41	Exercise Science	10 years ago in HS
Oscar	41	Biology	Recently in HS
Amir	43	Chemistry	No physics in HS
Evelyn	43	Behavioral Neuroscience	Recently in HS
Lauren	43	Behavioral Neuroscience	Recently in HS
Zion	43	Aerospace Engineering	Recently in university

during the course introduction speech, participation in the interviews was voluntary. Ultimately, there were six students from each group (i.e., experimental and control) who volunteered. As these students form the core of the qualitative data analysis, a summary of their pseudonyms and characteristics of interest is shown in Table I.

In many cases, students in nonphysics majors take the course at the very end of the degree in order to satisfy their degree requirements, by which point they may not have done any physics since their freshman year and thus not remember much of the material. Because of this possible asynchronicity and the aforementioned diversity of students, very few assumptions can be made about the physics background of the students in the course, which makes it difficult for a given curriculum to be suited to the levels of all the students.

To compensate for this, we decided to form maximally heterogeneous groups of three or four students based on students' gender, major, and year in the major. The intent of this approach was to capitalize on the inherent diversity of the students who enrolled in the course and allow their different strengths and weaknesses to complement each other, thereby improving group effectiveness. This is a perspective supported by the literature [39–41]. However, there are also arguments for having more homogeneous groups [42–44]. For a critical analysis of the approaches as well as of their arguments and implications, see the work of Esposito [45].

B. Modifications to the course for the experiment

We decided to create a labatorial-based curriculum for comparison with the traditional one. To that end, we designed six labatorials analogous to the existing traditional labs in terms of the core concepts and the experiments performed. Note that it was necessary to modify some of the core lab content because some traditional labs involved several concepts or experiments, thereby making it difficult

to adapt to a labatorial format. In many cases, this was because incorporating conceptual and other types of questions would make the labatorial too long (despite in-class lab reports no longer being required), and so we eliminated one of the experiments or concepts addressed to ensure that each labatorial was feasible and had a clear focus. The main experiments and concepts for each lab for both traditional labs and labatorials, which provide the basis for the learning outcomes that will be examined to address RQ2, are summarized in Table II. For a sample labatorial worksheet, see Supplemental Material [46].

Another important consideration in the design of the labatorials for the course was that labatorials are normally performed in conjunction with a corresponding lecture course. Namely, students should be exposed to the ideas in the lecture, from which perspective they refine the most challenging concept(s) in the labatorial. Were students to not have any preparation before a labatorial, they would likely not be able to participate in discussions with their teammates and thus not be able to get the most out of the labatorial. Since there was a high chance of students not taking the lab course at the same time as the corresponding lecture course, we introduced mandatory prereadings for the labatorial version of the course to act as a substitute for the lectures. The readings were comprised of selected webpages and/or selected sections from the lecture course textbook. In turn, we introduced mandatory summary writing, which was to be shown to the TA at the beginning of each lab.

After designing drafts of all six labatorial worksheets, each of three graduate students tested two of them, going through them independently and then providing feedback. We then made modifications based on a pilot study. Of the course sections that semester, only one, which contained 12 students, was run as a labatorial course. While interview data were collected during this phase, these results are not considered in the formal analysis process, and no formal

TABLE II. Core concept(s) and experiments(s) addressed in each lab.

Version	Lab number					
	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6
Traditional lab concept	Density of solids or liquids	Vector addition of forces	Spring constant	Centripetal force	Collisions and g	Pendulum period or energy
Traditional lab experiment	Using a pycnometer	Using a force table	Measuring force and period of a spring	Tension of a horizontally rotating object	Measuring coefficient of restitution and g	Verifying period formula and energy conservation
Labatorial concept	Density of solids or liquids	Vector addition of forces	Spring constant	Centripetal force	Collisions and g	Period of a pendulum
Labatorial experiment	Using a pycnometer	Using a force table	Measuring period of a spring	Tension in a pendulum	Measuring coefficient of restitution	Deriving period formula with dimensional analysis

analysis was performed at this stage. Based on experiences observing class sessions, the data collected during the pilot, and discussions with colleagues, certain changes were made in moving from the pilot to the full-blown study. In particular, it was decided to add weekly postlab quizzes for all the lab groups as well as postlab surveys for the TAs to complete.

C. Methods of data collection

1. Interviews

One of the main sources of data for investigating RQ1 and RQ2a was the set of semistructured student interviews. Student interviews took place between the first and second labs (the pre-interviews) since we could not meet the students until the first lab session, and after the sixth lab but before the final exam (the postinterviews). The timing of the postinterviews does not cause any issues for our analysis since the interview questions were focused on the research subquestions targeting general features of the labs (including how social interactions in the lab affect the student learning experience), not the one that is content based (how students' learning outcomes compare between the two lab approaches).

An important point regarding the design of the interview questions was that they could not always be the same for the labatorial students and the traditional lab students. However, because this is a comparative study, all differing labatorial student interview questions have a corresponding traditional student interview question that essentially targets the same point of interest. An example of this design choice can be seen in the question regarding the way students work in teams:

> *Labatorial*: Could you tell me what it felt like doing a lab like this [without a protocol] for the first time?
> *Traditional*: What do you think about the protocol format of labs?

However, some questions could be asked in the same way for both groups, such as the following question asking about their interactions in the lab:

> *Both*: Can you describe how you felt about your interactions with your partners and the TA throughout the session?

For the full labatorial and traditional lab student interview guides, see Supplemental Material [46].

Semistructured TA interviews were also conducted to complement the student interviews. However, the main differences are that the TAs were only interviewed at the end of the course and that their interviews are being used as data for triangulation rather than as a primary data source. It is also worth noting that while there was a TA present for each of the five lab sections, there were only three distinct TAs; two TAs (pseudonyms Isaac and Liam) each overlooked one labatorial section and one traditional lab section, and one (pseudonym Justin) was only responsible for a labatorial section. This means that Justin could not

necessarily comment on experiences about labatorials and traditional labs for our course, limiting his responses to a degree. However, we believe that the inclusion of his interview material is nevertheless valid since he had experience in a different, traditionally run introductory lab course during the same semester, and thus he could still speak to the similarities and differences between labatorials and traditional labs.

2. Additional qualitative data sources

We also qualitatively examine the 12 interviewees' responses to selected post-test and final exam questions. These will form the foundation of our investigation of RQ2b, which involves content-specific learning objectives. We also examine the responses of six other students from each group who are selected in order to ensure that we examine the results of a diversity of students in terms of their overall final exam grade, gender, and major. This is important for increasing the reliability of the results as it helps establish whether or not our interviewees are representative of the whole class.

Some other qualitative data sources are also considered for the purposes of triangulating with both the interviews and examination responses. First, this includes observations as a passive observer, i.e., an observer who does not interact directly with the participants; one of the authors (F.L.) attended the lab sessions of the different course sections. For a sample set of observations, see Supplemental Material [46].

In addition, the TAs were asked to fill out qualitative surveys at the end of each lab regarding students' understanding of specific concepts targeted in the lab. For a completed example of such a survey, see Supplemental Material [46]. For a labatorial, they would complete the survey immediately at the end of the class, while for a traditional lab, they would do so after grading the students' lab reports.

Lastly, students' writing products are also examined. For labatorials, this refers to the labatorial worksheets, while for traditional labs, this refers to the lab reports with particular focus on the discussion and conclusion sections since this is where concepts are the most likely to be mentioned considering the report template they are instructed to follow.

3. Quantitative course assessments

First, we administered a general pretest to all the students at the beginning of the course. This pre-test serves not as a base for pre-post comparison, but rather as a means of establishing equivalence of the two groups, which would then allow us to meaningfully compare them. As such, there is no corresponding post-test for this general pretest. Because this is a mechanics lab course, we compiled six questions from the Force Concept Inventory (FCI) [47] that span some key topics in kinematics and dynamics.

Research Question	Sub-Question	Student Interviews	TA Interviews	Observations	TA Surveys	Worksheets/ Reports	Quantitative	Final/ Post-Tests
1	A	✓	✓					
	B	✓	✓	✓				
2	A	✓	✓	✓	✓			
	B			✓	✓	✓	✓	✓

🟧 Comparison of general features of labatorials and traditional labs

🟦 Analysis of content-based learning outcomes

FIG. 1. Structure of research questions and subquestions for triangulation.

However, these questions are only used for the purpose of this general pretest and will not be used again once equivalence of the treatment groups is established. For the questions included in this general pretest, see the "Conceptual Quiz on Classical Mechanics" dated May 7–9, 2019 found in the Supplemental Material [46].

Next, because we wish to compare students' conceptual learning outcomes in labatorials and traditional labs to answer RQ2b, we examined student scores collected throughout the course. Since we do not perform any pre-post testing, the sets of scores for the two groups are analyzed and compared purely in post.

The first of these are the conceptual post-tests (postlab quizzes) containing one to two questions that were designed for each topic and administered after each lab. For a sample post-test and the rubric used to grade it, see Supplemental Material [46]. While there were preexisting weekly pretests for the traditional labs (unrelated to the aforementioned general pretest), we could not perform pre-post comparisons since equivalent weekly labatorial pre-tests could not be designed. This was due to logistical constraints as well as the differences in lab content for each group.

In addition to the post-tests, we examined the questions from the final exam of the course. In part due to logistical constraints, we did not completely modify the original final exam. However, we designed six conceptual questions specifically targeting the core concepts of each lab and introduced them into the final exam. (These questions are unrelated to the six FCI questions utilized for the general pretest.) Furthermore, some of the original questions that were either not conceptually interesting or simply redundant with the new questions were removed.

Cross tabulating the data sources versus the research subquestions as shown in Fig. 1 allows us to visualize which data sources are pertinent to which subquestions, which will be helpful for triangulation.

D. Methods of data analysis

1. Interview analysis

When qualitatively coding interviews, one typically examines each interview with a certain approach in mind. Namely, after dividing an interview into key segments, one derives codes associated to key passages or ideas in these segments, all the while comparing them with the interview as a whole and the other interviews examined so far. One then iteratively groups together redundant codes so as to ultimately derive five to seven core themes or categories. Upon coding all the interviews, the ideas that are common across all interview transcripts are then considered general by assumption [48].

Because this study is explorative in nature, we begin by applying an inductive coding procedure on the student pre-interviews. This allows us to identify ideas in the transcripts of both labatorial students and traditional students with as few preconceptions as possible, allowing us to compare the results with the literature and present a minimally biased comparison. After examining the transcripts of all 12 interviewees, we derive a set of themes and major categories that we thought would not change greatly with the postinterviews. While new ideas could most certainly still emerge in those interviews, our code structures are created such that existing categories and themes can simply be added to or modified should a segment of the postinterviews not naturally fall into an existing bin. As such, we turn to a predominantly deductive approach for analyzing the postinterviews.

We believe that a combination of approaches as described is appropriate when investigating a phenomenon that is already documented fairly thoroughly in the literature. The inductive aspect allows us to derive themes in a minimally biased way, and the deductive aspect allows us to further build our comparison as well as target our research questions more purposefully.

When moving on to the TA interviews, we again apply an inductive coding procedure since that data are used for triangulating with the student interviews, which form the core of the data for addressing the research subquestions pertaining to general characteristics of labatorials and traditional labs (RQ1a, RQ1b, and RQ2a). As such, it is necessary to minimize the biases present in the themes derived and verify if the comments of the TAs are consistent with those of the students. However, since we are prioritizing the student interviews, we frame the remainder of the discussion on qualitative coding with regard to the students.

2. Specific coding process

Since the goal is fundamentally to compare labatorials and traditional labs across various dimensions, it seemed to be most fruitful to produce two primary sets of codes: one for the labatorial students and one for the traditional lab students. However, we did not formally distinguish between codes that originated from pre-interviews and those that originated from postinterviews. Once all the interview transcriptions were completed and verified for accuracy by the students, we began coding the pre-interviews in the same order that they were transcribed. This coding process continued until all the labatorial student pre-interviews were analyzed. Before moving on to the traditional student pre-interviews, the labatorial categories were reorganized since some of the main themes had already begun to emerge. This whole process was repeated for the traditional lab students, building up a new set of codes and categories and ending that portion in a fashion similar to that just described for the labatorial students. With a set of probable main themes established for both student groups, we then began deductively coding the postinterviews similarly to the overall process just described.

While the coding process for each set of pre-interviews always began inductively, it was impossible to completely eliminate bias in the way we think about codes and themes. Even before collecting any data, one codes interviews with the aim of addressing certain research questions. In our case, the primary motivation for the codes created and themes derived originated from the informal examination of the student interview transcripts from the pilot study. Even without formal analysis, a notion of "support" resonated throughout the interviews, which was in large part associated to their peers and the TA. This appeared particularly pertinent to addressing RQ1 on the student learning experience. However, these particular manifestations of support also bore a strong resemblance to the idea of scaffolding previously discussed, which play an important role in building conceptual understanding. As such, although students' learning experience and their conceptual understanding are two distinct aspects of the research, the author (F. L.) aimed to understand both as being the result of a unified set of underlying mechanisms. Therefore, the

notion of support was kept in the back of F. L.'s mind as codes emerged throughout the interview analysis, considering any new codes and categories as potential means of formalizing the notion of support in terms of mechanisms of scaffolding.

Lastly, note that due to the number of different aspects of labatorials and traditional labs that are addressed in the interviews, it can be difficult to isolate which differences between the approaches might have led to an observed result. This is indeed a concern that cannot be ignored because many of these aspects are interrelated. However, because our interview questions consider one aspect of the labs at a time (e.g., interactions with peers, interactions with TA, structure of the lab manual, etc.), it becomes possible to associate the code corresponding to the aspect addressed in a given response to the impact(s) it had in each type of lab. While there indeed may be multiple factors contributing to a given observed result, this is a nonconcern, as we merely wish to identify the codes and themes pertinent to answering the research questions for each lab group. A detailed analysis of which factor had the largest impact on a given outcome is beyond the scope of this manuscript.

3. Analysis of assessments

In analyzing student assessments, we are most interested in examining the individual responses for the post-test and final exam questions rather than overall exam averages. This is because we wish to compare students' learning outcomes pertaining to specific course concepts in order to answer RQ2b. Since the final exam contained several questions that were not conceptual in nature or were in a format not conducive to inferring conceptual understanding such as factual multiple-choice and short-answer questions, an average alone would not yield information pertinent to our research question. Therefore, the focal point of the analysis will remain qualitative in nature, although some statistical analysis will be used to preliminarily explore the data, conjecture regarding patterns in the data, and direct the primary, qualitative analysis. Students' question scores will also play a role in triangulating with the other data sources pertinent to RQ2b, but this process too will be largely qualitative.

Based on the improved conceptual understanding of students documented in the literature [33,36,37], we would expect to observe a statistically significant difference in overall performance between the labatorial and traditional lab groups. However, since we only wish to examine students' understanding of specific concepts, we will simply use this prior result to motivate a heuristic for identifying possible questions of interest among the assessments. Namely, we will first tentatively identify which questions exhibit significant differences in performance using student's independent samples t test at the 95% confidence level to compare the mean score for each question for the labatorial group and the traditional lab group. This

will also give some preliminary insight into the types of questions students from each group perform better on. We believe this heuristic to be appropriate in this case since the questions were designed by F. L. to assess specific learning objectives. As such, a statistical difference in performance would serve as an indicator of a possible difference in conceptual understanding. Once notable questions are identified, we will hypothesize a general pattern regarding which types of questions are typically performed better on by labatorial and traditional labs students and then examine students' written responses in detail. Note that for a postinstruction comparison between two groups such as the one proposed to be meaningful, the equivalence of the groups must be established, which is why we will precede the aforementioned analysis with the general test for equivalence described in Sec. II C 3.

In addition, note that we acknowledge the limitations of the t test, both in terms of its limited practical significance and its appropriateness given the assumptions that must be met for it to yield meaningful results. The results of the labatorial and traditional labs groups are certainly independent. Furthermore, although homogeneity of variance may not always be satisfied, the analysis software used by the authors (SPSS) can detect this and automatically apply Welch's t test instead, which typically yields nearly equivalent p values but has slightly lower power than student's t test [49]. However, the question score distributions may not always be normally distributed. To avoid this difficulty, we also perform a Mann-Whitney U test [50], which does not rely on a normality assumption, to cross-check with the t-test results. Nevertheless, these back-of-the-envelope tests serve only as a guiding heuristic in our case, not as evidence from which to draw any conclusions, and so they will not impact the results of our primary analysis.

III. RESULTS

In order to structure the discussion of the most poignant themes that emerged from the student interviews, we present a summary of the main themes in Table III. The table is structured so as to compare and contrast the themes for labatorials and traditional labs with regard to students' learning experience and positive and negative affects on

their conceptual learning. This allows us to directly address RQ1 as a whole and RQ2a, respectively. Each theme will then be discussed in turn within the context of the pertinent research question. Although there is some overlap across research questions for certain themes, this does not invalidate the relationship established between any given theme-research question pair. The specific relationships between the themes and research questions are illustrated using the concept map in Fig. 2, where similar themes are condensed into a single node. Following the discussion of the themes, the analysis of students' course assessments will be presented.

A. The student learning experience

We first examine RQ1:

How can the learning experience differ between labatorials and traditional labs?

By "learning experience," we refer to students' collective affective experience in the lab. In particular, we are interested in investigating (i) the role of social interactions in the lab on students' experience and (ii) what makes a lab course satisfying to students, corresponding to RQ1a and RQ1b, respectively. After analyzing all the student interviews, we found that there is a unifying theme that can be used to mostly address these first two points. Therefore, we will begin by discussing this theme in detail for labatorials. We will then discuss the corresponding themes and address the research question in the context of traditional labs.

The idea of scaffolding is the highest-level theme that emerged from the labatorial student interviews, which is in accordance with our main hypothesis. In particular, in Sec. II D 2 we discussed the affective element of scaffolding, which we referred to as "support." A variant of the theme of support also emerged from the traditional lab student interviews, and it is this theme and its subthemes or codes (as listed in the first column of Table III) that are essential for addressing the two aspects of our research question. What is critical to consider is that while there are some broad similarities between the codes used to capture the themes for each group, the forms of support manifest differently in each type of lab.

TABLE III. Summary of themes from student interviews.

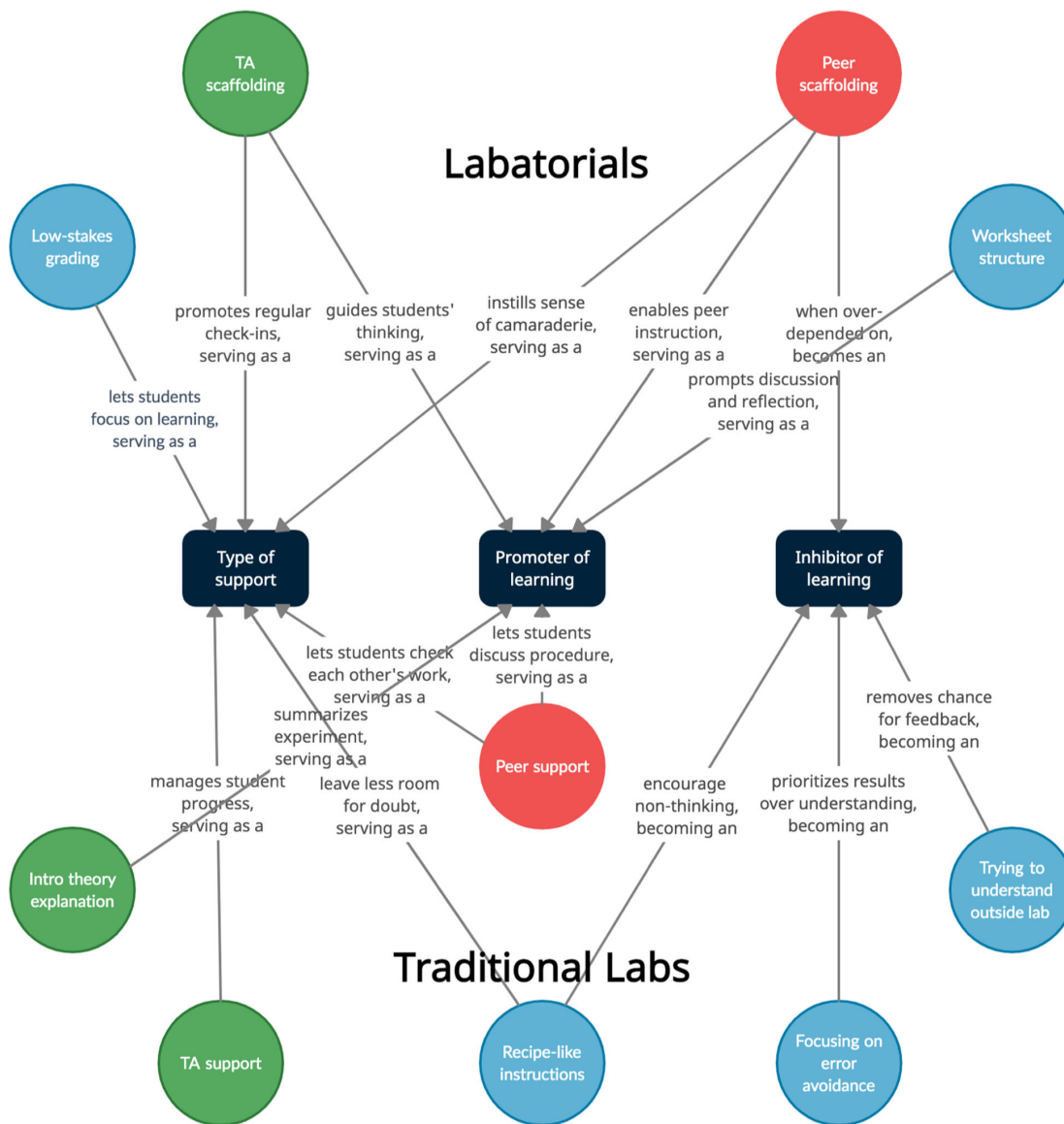| Lab type	Dimension of research questions		
	Support types	Promoters of learning	Inhibitors of learning
Labatorial	Peer scaffolding	Peer instruction	Peer over dependence
	TA scaffolding	TA scaffolding	
	Support due to low-stakes grading	Labatorial structure	
Traditional	Peer support	Peer interactions	Focus on error avoidance
	TA support	Intro theory explanation	Recipelike instructions
	Support due to explicit lab instructions		Trying to understand after lab is already done

FIG. 2. A concept map summarizing the key relationships between the student interview themes and the dimensions of the research questions. Red, green, and blue nodes indicate overarching themes broadly related to peer interactions, TA interactions, and lab structure, respectively. Each arrow corresponds one of the specific themes listed in Table III.

1. Labatorials

In labatorials, peer support was very important for creating a positive experience for students. As suggested in student Catherine's postinterviews, many students perceived a sense of camaraderie in their teamwork. Beyond simply working together for the purpose of completing the lab, they found reassurance in knowing that they were going through similar struggles, helping them feel comfortable in sharing their doubts and relatively relaxed despite the challenges of the labatorials. When asked what was most special about the teamwork in labatorials, Derek stated that

> …it's the, I don't know if it's the right word, but the camaraderie of it. Because by myself in labs, if I get

frustrated, I just feel so lost. You kind of feel like, Oh god, everyone else maybe gets it and I don't. What's wrong with me? But when you're in a group it's more like, Oh, they don't understand either! Or I know more about some things and they know more about some things. It's a lot more just relaxing, easier to just focus on the lab itself. And it's good to bounce off ideas and ask for help from your partners. So I'd say it's just a better experience, just less stressful like in terms of being by yourself.

Jessica also strongly resonated with this idea, additionally commenting that the team's mutual dependence instills a sense of group accountability as individual accountability:

So when you're doing it with the misery of… Not that we were miserable, but if you're going through something that's challenging, it's much easier to do when you're with other people who are in the same boat and feeling the same stresses and things like that. And that they're rooting for you, because their success also depends on you, whereas in other labs it's like, Screw you, I'm just going to do my own thing and you can go ahead and drown.

Interestingly, many interviewees indicated that this sense of camaraderie extended beyond their peer interactions to their interactions with the TA as well. Jessica stated that she "really liked, like not working with him, but how he came through" whenever they were struggling. Catherine additionally expressed that "the nice thing about [the lab] is that even with the TA […] it felt like a team effort toward understanding." This suggests that the TA relationship was one of collaboration as much as one of guidance. Additionally, all students expressed that they felt that the TA was always very involved, regularly checking in on students. This helped Catherine to "really feel supported by the TA."

The checkpoints in the labatorials in particular helped encourage students to share their doubts with the TA and feel supported by him, particularly due to the checkpoints being a core element of the lab itself. Namely, Emma expressed that

…it's expected to ask questions, and it's expected to get the go-ahead before moving on. I like that so much more than just feeling like you're on your own if you don't understand it because you haven't prepared properly. I like having the help available, and that you're expected to use the resources and help of somebody there.

The checkpoints additionally helped alleviate students' stress in the lab by helping them feel confident in progressing through the worksheet. As expressed in part by Stacy, this is because "if you made a mistake, the TA will tell you right away. You don't have to keep going and then figure out [later on] that you screwed up." However, being a relatively strong student in the course, Stacy also experienced a certain drawback with this form of TA support. Often reaching a checkpoint and already understanding the material quite well, she "[had] to wait for the TA to come" despite feeling ready to move on. Even so, she had no other qualms, understanding that "it's more for [her team] to make sure [they're] on the correct track and try and interact together."

Related to the TA's support was the support incurred by the grading scheme of the labatorials. Because the grading scheme for labatorials is largely participation based, students did not have to worry about making mistakes in the lab, which helped them feel more comfortable expressing their doubts. (For the exact grading scheme, see the Supplemental Material [46].) As expressed by Emma, "you're not being tested on if your assumption is right at all, but you're kind of asked to think about it beforehand and test it out. So there's no penalty in making a guess." Jessica also echoes this sentiment, indicating that not feeling any pressure due to grades also benefited her team, helping them interact more smoothly:

Here because we're working together in a way that it's okay to ask questions, it's okay to not get the answer completely right. As long as you understand why it's not right it's okay… Even that was more relaxed. It was easier to laugh and learn instead of feeling like crap because you're the teammate at the bottom or like, Okay, I have to drag these two.

As a result of the various forms of support present in labatorials, some students also underwent changes in their perspectives on and feelings toward physics in addition to having a satisfying learning experience. Catherine in particular, whose feelings about physics before the course were bleak, underwent a major transformation in outlook. She now "that [she] feels better about moving forward in physics in general," and she and Jessica both expressed that physics became less scary to them by the end of the course. Additionally, Stacy said that she came to better appreciate the importance of "not just accepting the equations as they are, but trying to figure out what each letter is doing, what each component is doing," suggesting a shift toward expertlike thinking. In addition, many students perceived a sense of relevance in the material due to the connections made with the real world. In particular, Emma stated that the experiments, which were posed as real-world problems, were "relatable to other areas that you're going to study in the future or presently," making the experience "much more satisfying" and motivating for students.

In summary, students' interactions between both peers and the TA in labatorials served as a form of scaffolding that was affectively perceived as mechanisms of support. Although the constant scaffolding (in particular by the TA) could be perceived as restrictive by stronger students, most students indicated that the overall lack of stress and sense of ease incurred by such support made for a more enjoyable overall learning experience, with many additionally undergoing positive shifts in their perspectives on physics. This is reinforced by several unprompted vocalizations of satisfaction by all the interviewees regarding the labatorial style. These results are consistent with the literature [19–23]. Such a positive lab experience also suggests benefits for students' conceptual learning. The ways in which the aforementioned forms of support served as conceptual scaffolding mechanisms will be thoroughly discussed following the discussion of the learning experience in traditional labs.

2. Traditional labs

In traditional labs, peer support and TA support were also themes that reverberated across several interviews. However, the ways in which these themes manifested in the lab and the role they played for students are different than they were for the labatorials and may not have been effective as scaffolding mechanisms. In particular, the group aspect of the lab was not as fundamental a part of the lab experience as it was for labatorials. While there were labs where students needed to work in teams, the interviewees appeared to often not work in a very unified way. Students would at times proceed at a similar pace, but take their own approach to different steps of the experiment. While not an experience from this course, Catherine expressed that in many of her past nonphysics labs that were run under a similar traditional format, although "[they] had] all read the same lab [...], [they] kind of [came] at it with a different approach, or [they didn't] know who [was] going to do what." As such, the interviewees' group work experience was often more akin to working individually in their groups rather than collaboratively driven. As expressed by Lauren,

> Since we didn't know our teammates very well, we decided to kind of go alone but share with them. So we'd do things on our own and share the process. So we'd do every step, write the results next to it, and then keep going, and then ask if everyone's ok with it.

This is consistent with most interviewees expressing that they prefer depending only on themselves. Some students do appreciate the value of teamwork, with Amir in particular stating that "[he] would love working with a teammate and having a discussion and working together, so long as their work ethic and personality is aligned." However, Adrian said that "[he enjoys] working alone, and if [he has] questions, [he'll] go up to someone and ask them," and most of the interviewees were of this mindset. He and most of the other interviewees expressed that they would ask each other "[questions] like, How did you set this up? or, What value did you get for this? Just to make sure [they] were on the same track in terms of what they were doing." Therefore, through being able to verify each other's work, peers appear to serve as a source of support for many students.

The TA in the traditional labs served as another source of support for similar reasons. While students would ask their peers first if they were not sure about something, they felt comfortable asking the TA questions if they needed additional help. Evelyn said "he seemed to be receptive to the sort of questions that [she] had," with Amir adding that "he's always willing to help [them] figure out things in a way that works for [them]." As with the questions peers would ask each other, the support provided by the TA typically involved clarifications or verifications.

Furthermore, the TA would always answer students' questions very directly, "without being hesitant or questioning what's going on," as phrased by Zion.

While the students all appreciated this support by the TA, many also indicated that the TA was typically not deeply involved with students involved throughout the lab, taking instead a more managerial approach. As described by Zion,

> He was [...] going around, looking everywhere, seeing how students are doing and all that, but he wasn't interacting with the students. He was just looking, and if he saw something wrong, he would say, Well this is wrong, you should probably not do it this way, you should do it that way. And that's it, that's all he did. He didn't really do much other than that.

Students nevertheless appreciated this managerial TA style as it reassured them that they were proceeding correctly, with students including Lauren feeling that he was "always very involved and very helpful" in this regard.

As Lauren also indicates, the main source of stress experienced by students in traditional labs was that incurred by grades. Because students want to score as well as possible, they want to follow the instructions as closely as possible and constantly verify that they are not making any mistakes. Adrian also added that "if your percent error is off by more than 10% [when checking in with the TA], you'd have to redo everything," expressing frustration at this possibility. As a result, the highly detailed nature of the lab instructions for the traditional lab experiments acted as a form of procedural support for the students; with instructions that are overall clear and easy to follow, there is a lower likelihood of making errors. Additionally, most of the interviewed traditional lab students were new to physics labs. As such, some expressed that having the detailed, step-by-step instructions helped explicate the basics of conduct in physics, thereby providing a source of guidance to beginners in physics. In particular, Evelyn stated that

> ...for the material that we're doing, it's sort of the only way that makes sense. Not that I know any other teaching methods, but it makes a lot of sense to have it that way just because for some people, maybe it is the first time they're in a scientific lab for these things. So it does make sense to sort of become acquainted with physics labs and labs in general. It does make sense to sort of spoon feed the material, or the experiment, I mean.

However, Oscar felt that there was a lack of guidance at times with regard to both the TA and the lab manual. Namely, he expressed that

> There should be much much more guidance, much much more. There should be like, We should do that and that and that, and if you have any questions just ask more

and more, you know? More guidance in, even the manual should have much more guidance in the labs.

Catherine's comments about her past lab experiences also complement Oscar's frustration, adding that "if you don't know to ask the questions leading up to that moment, you're docked a bunch of points, and you don't really understand what you're doing because nobody's guiding you." Ultimately, we can see that many students, including Evelyn, feel that "to get the result that you want, […] you do need that sort of support in terms of the procedure and the instructions." Therefore, although most of the interviewees appeared to be satisfied with the support present in the labs since they could get help whenever they needed it and felt at ease doing so, the need to get a good grade in the lab may still be a large motivator and source of stress for some students. The impact that this mindset may be having on their learning will be discussed in the following subsection.

Regardless of the stress, some students did mention additional positive aspects of their learning experience. For example, Lauren said that seeing the theory applied in the lab helped "bring [her] back to reality and see how it is, remembering that [real-world] connection." In addition, one student (Adrian) indicated that his outlook on physics changed since taking the course, expressing that "physics is everywhere. Like it exists in everything that we do, […] it exists outside of the lab as well." On the other hand, some students like Evelyn, for example, were more apathetic regarding the impact of the course, sharing that the labs were "a little bit boring" and that "as much as [she] wanted to have a takeaway from it… It was [just] the information that it was supposed to give [her]," suggesting an unsatisfying experience.

In summary, with the exception of Adrian, who expressed that the lab "wasn't as bad as [he] thought it was going to be," none of the interviewees vocalized satisfaction with the course without the prompt of a question, albeit their learning experience was not negative as is often reported anecdotally and in the literature [14–16]. Nevertheless, there were certain forms of support present in traditional labs, namely, the procedural support provided by the lab manual and that provided by peers and the TA in acting as a resource for students to voice their doubts and verify their work. These often helped students partially cope with the stress in the lab and thus feel more at ease, although some students felt that the lab experience was not scaffolded enough. The degree to which the aforementioned forms of support acted as mechanisms of scaffolding for conceptual learning in traditional labs will be discussed following the discussion on students' conceptual learning in labatorials.

B. Conceptual learning in the lab

We next examine RQ2:

In what ways do labatorials and traditional labs promote the development of conceptual understanding?

By "conceptual understanding," we refer to students' mastery of the physics concepts underlying the phenomena under investigation, beyond just being able to perform calculations or take measurements. In examining the student interviews, we are interested primarily in the general promoters and inhibitors of conceptual learning in each type of lab, which are pertinent to answering RQ2a in particular. We focus on conceptual learning in general for the remainder of this subsection, structuring the discussion according to the themes listed in Table III and the relationships illustrated in Fig. 2 as with RQ1. Furthermore, most of the main themes pertaining to labatorials and RQ2a were found to directly relate to the constructivist idea of scaffolding, whereas the traditional lab themes mostly exhibited a lack of scaffolding. Note that students in both groups expressed how hands-on work helped them make connections with the real world and better understand the lab concepts. However, this is unrelated to scaffolding, and so this theme will not be discussed here.

1. Labatorials

For labatorial students, the forms of scaffolding inherent to the design of labatorials all played a significant role in students' conceptual growth. Consistent with students' comments regarding their affective learning experience, peer instruction was an important mechanism assisting students' learning. This occurs in part because of the reciprocal nature of the process. In particular, Derek noted that he and his teammates did not simply provide each other with the solution to a problem or the answer to a question, but rather frequently engaged in discussion, saying that

It wasn't just like, Here is the equation that you use for this. Solve for this. […] That's what was really helpful for the groups. It's that back and forth talking about which direction we should take it.

As such, the students in labatorials were deeply cognitively engaged with the concepts of the lab, discussing until they all came to the same answer. This type of collaboration was also encouraged by the design of the labatorial questions themselves, which were challenging enough that students usually needed to collaborate if they wanted to proceed. Therefore, students became more engaged in the problem solving process while simultaneously developing their ability to think independently.

The highly supportive nature of the teamwork in labatorials also encouraged more meaningful discussion between peers, often strengthened by the heterogeneity of the teams. Catherine perceived both of these benefits, saying that

To feel like you're in a safe enough space that it's like, Hey, I actually don't know what I'm doing. Could you explain to me why you understand this? We all had

moments like that, where one of us was the one who knew more, and the other was the one who knew less. We were even, but we all came out more knowledgeable.

The proficiency differences inherent in heterogeneous grouping schemes bring about both a strength and a weakness of heavily collaborative learning frameworks. As stated by Jessica, "it's good to learn [in a team] since when someone has a deeper level of knowledge, it's easier for them to teach others." Furthermore, because students need to "understand how to tackle explaining […] at different levels," as phrased by Stacy, students at all levels also learn through the act of teaching. In particular, Emma, a weaker student, expressed that "[they] started to understand a little bit better each other's learning styles, so [they] adapted to each other better," feeling that her teammates "were very very patient with [her]." However, while Emma put in a great deal of effort despite struggling in the course, there is nevertheless the risk of the weaker students in the group overdepending on the stronger students for proceeding through the lab. Aside from the case where a student is unwilling to cooperate, which can occur in any type of lab, a student may not have sufficient time to ponder on the concepts, even when they are explained by their peers. This pace matching of students' work is inherent to the structure of the labatorials, imposed via the checkpoints, and so the associated difficulty may be mitigated by reducing the amount of content in the worksheet.

Nevertheless, the labatorial worksheet structure, in conjunction with the TA's guidance, served as an important form of scaffolding for helping students learn. In particular, the strategic interventions of the TA at the checkpoints helped guide students to an understanding of the concepts while ensuring that all students came to the same conclusion and did not build on misconceptions. As expressed by Catherine,

I'm seeing step by step kind of what's happening, and the fact that [the TA] would come over and check in with us each time, and kind of affirm, Yes, you're doing this right, or, Actually, why don't we think about it this way instead? is very helpful. It makes me feel like, Yeah ok, this makes a lot of sense.

The overall design of the worksheet, which aimed to build up students' understanding of the concepts question by question, also helped students understand. Jessica "felt how the manual was structured and how it progressed throughout," which helped her gradually construct her understanding. Stacy also "really liked the way that [the worksheets] followed teaching something that was very important, but without really [saying], that's just how it works. It was kind of like, figure it out by yourself how it works," which helped her fully grasp the concepts by the end of the lab.

This was also enhanced by the inclusion of explicit conceptual questions. Stacy stated that they "[made] sure

you were actually thinking, critically thinking about things and not just accepting [them]" as well as "[helped] set up the brain to keep going [through the worksheet]." In addition, many students including Jessica felt that the questions "helped [them] later for the actual experiment. The conceptual problems changed how [they] were actually looking at the experiment [compared to] when [they had] just read [the instructions]." Emma indicated that such a lab format, which allows students to tackle concepts in tandem with the hands-on work rather than try and figure things out after the lab, was helpful for her conceptual development with regard to dealing with misconceptions:

It's because it's hands-on and it's also theory, and actual data gathering at the same time. So you have to apply what you know and what you're learning reciprocally to each other. So you have kind of two chances to correct what you wrongly thought, I guess.

Many students also indicated that the inclusion of conceptual questions helped with problem solving and understanding the mathematical formalism associated to the concepts. Therefore, the conceptual questions played a large role in deepening students' engagement with the concepts as well as the experiments of the labs, encouraging them to not passively proceed without understanding.

The prediction questions of the labatorials were also important for developing students' conceptual understanding. According to Quincy,

All the predictions, they just helped you write out and then discuss with your teammates about your own ideas, about that event, about that problem. And then when you go through the lab, they start to change because not exactly everything you think is right. After you go through the lab, you go through all the experiments and all the work, and you will find the final result of that problem. And then you will understand that, Ok, I was thinking wrong at first, and now I need to think in that way for things to make more sense.

Quincy's description of this learning process in the lab, which most of the interviewees expressed in one form or another, is exactly consistent with the elicit-confront-resolve sequence for tackling misconceptions described by McDermott *et al.* [50] in the context of the Physics by Inquiry curriculum on which the approach of labatorials is intrinsically based; the worksheet elicits students' preunderstanding of the concept, allows students to confront their misconception(s) through the experiment, and then resolve the conflict(s) in their understanding through discussion. It is the extensive scaffolding inherent in labatorials that allows this development of conceptual understanding to occur.

In summary, the integration of diverse mechanisms of scaffolding in labatorials provided numerous benefits for

students' conceptual development. Although the risk of over-depending on one's peers exists due to the extensively collaborative nature of labatorials, peer instruction was nevertheless a powerful means to assist learning. The additional affective benefits of such scaffolding (as well as those due the overall absence of pressure due to grading) also allowed students to focus more on their learning in the lab, which, in conjunction with the inherently scaffolded worksheet structure, improved their overall engagement with their peers and the concepts in the lab. Furthermore, although labatorials emphasize understanding concepts over developing experimental skills in this way by design, such an approach also proved to be beneficial to students' understanding of the experiments. By virtue of these pedagogical advantages and the aforementioned forms of support inherent to the labatorial approach, all of the interviewees "think very highly of it overall" and expressed a preference for labatorials over their past traditional labs, as with Emma; Quincy stated that he "[thinks they] can learn more from it than the traditional [labs]," and Catherine said that she "would choose this over a traditional lab every time."

2. Traditional labs

In traditional labs, there was also some scaffolding present due to the interactions between peers, which had a positive impact on their learning. However, such scaffolding was limited in scope. As previously mentioned, these interactions typically involved students checking each other's results. Furthermore, Adrian, among others, described a divide-and-conquer style of group work wherein students would subdivide tasks and assimilate their progress for the sake of efficiency. Namely, "[his peer] did one thing, [and he] did another thing. [They] kind of just worked together" without thinking through each task together. Therefore, the group work in traditional labs was typically not very collaborative. Nevertheless, students did also at times explain things to struggling peers. Even during the individual labs, Adrian stated that he "would just ask [his] neighbor doing the same experiment and then see if they're having the same issue, or if they could give [him] an explanation or help [him] out," and Amir said that "if there [were] enough people [with the same issue] and one person says, I don't have any experience with this, then [they] created] kind of like a little side gig where, Let me show you how, or Is anyone confused with this?"

Zion indicated that he understands the conceptual benefit of discussion in the lab. He and his teammates would "communicate to one another. [They would] say, Why is this happening? Why is that happening? so [they] understand what's going on. Because sometimes, even following the procedure, it doesn't mean [they] understand the results." He also indicates that there is the potential for very fruitful collaboration in traditional labs, albeit it often does not occur unless the students take the initiative:

If, for instance, there's a problem going on and somebody asks this question and goes against the idea of what's going on like, I don't think this is okay to do, maybe we should do it another way, like this kind of, not attack, but confrontation about the idea is beneficial for the team, beneficial for the experiment, and that would be ideal.

Therefore, we see that, as with labatorials, students in traditional labs can scaffold each other's understanding via discussion when there are students with different levels of knowledge or when team members think critically about the experiment and the underlying concepts. However, most student responses indicate that the discussions that took place between students largely did not exceed the procedural aspects of the lab. This may be due in part to traditional lab students' focus on error avoidance, as previously introduced, which may be acting as an inhibitor to conceptual growth. While experimental accuracy and precision are important from an experimental standpoint, it can lead students to focus solely on following the lab instructions as closely as possible and not think through the core conceptual ideas as they work through the lab.

Furthermore, because there is a grade associated to these criteria, there exists a pressure due to grades that can cause students to not be focused on the learning in the lab. There are often students like Zion who typically do care about understanding what they are doing in the lab, including understanding the concepts. Oscar, for example, felt that avoiding experimental errors was also important for conceptual understanding since "if you do it in a wrong way, you'll understand a really really wrong principle." However, the recipelike, correctness-focused lab structure was generally less conducive to conceptual learning in the lab, which is well explicated by Zion's comment:

You would have to kind of do a little extra work if you really want to understand it. And if you don't, then you're going to go in, you're going to follow a bunch of steps, […] and then that's it.

A related inhibitor to conceptual learning is that because students may go through the motions of the experiment without thinking through the concepts deeply, they may not address any conceptual difficulties while in the lab. In most traditional lab courses, where lab reports are written at home, submitted a week later, and then returned the following week, students will often only begin trying to understand the ideas of the lab after the lab has already ended. Additionally, they only receive feedback on their understanding when their report is returned to them, which limits the possibility of difficulties being addressed while the ideas are still fresh, as well as that of productive feedback being given by the TA. This is in direct contrast with the instant feedback possible in labatorials, where

students work through all the concepts in the allotted class time. Although the lab reports for this traditional lab course were written at the end of class, many labatorial interviewees addressed such issues with regard to their prior traditional lab experiences. Catherine, for example, expressed that you may not thoroughly understand what is going on during the lab, and so you "just have to figure out what you did and why you did what you did since you don't know going into it why you were doing half the steps." Stacy additionally compared the value of thinking through the concepts in the lab to that of doing so at home, further illustrating why students likely will not confront their misconceptions through traditional labs:

> I think [traditional labs are] less about thinking, and more about if you can follow steps. [...] Sometimes we just try to rush and do whatever we can to get results, and then we get home and deal with it. In class, you have the opportunity to talk with your lab mates, and the TA trying to figure out [a problem] on the spot, and sometimes that's more valuable than trying to Google it once you get home.

Nevertheless, one element of the lab that played a role in helping students begin thinking more about the concepts was the TA's introductory theory explanation for each lab. This explanation would summarize all the essential theory for the experiment as well as the core experimental steps and apparatuses. Adrian expressed that "having the little discussion at the beginning, clarifying and kind of information" helped prime his mind before "seeing [the experiment] live and in action and actually performing it." Evelyn also added that it "[helped] to know sort of what the actual essence [was] if there's something [she] didn't understand and things like that." Therefore, the introductory theory explanation can serve as a base for students to more deeply understand the experiment and the related concepts. However, because it does not directly address common misconceptions and does not scaffold students' thinking, it will not help them overcome cognitive dissonance. Furthermore, although students may feel that they understand something by reading an explanation, this is often not actually the case. The analysis of students' exam responses, which can be used to address this point for the key course concepts, will be discussed in Sec. IV C.

Incidentally, Amir expressed that the learning experience in the lab was not scaffolded enough overall, despite the presence of the introductory explanations. In particular, the recipelike lab instructions of the labs made him unmotivated to learn:

> I found when I started off, ok just memorize everything in the lab manual. Read it, memorize it, memorize it. And then I just stopped kind of… Like I'd read it, take a few notes, make sure I see [the basics]. But that doesn't

really encourage, you know? It kind of just makes you… Makes me go complacent, apathetic.

This sentiment extended to the lab reports for the course, saying that he "found [the format] really formulaic. And [he understood] where it was kind of like, Data. Conclusion. Error. Discussion. But [he thinks] if there was a way where it started off and tapered," he would be able to learn more from it. "[He's] not saying [they] should be spoonfed through it, but [he's] saying there could be a lot more of, Do you understand this concept? [Talking] about it and [asking] very specific questions." Additionally, "[he thinks] if it were more kind of like a guided approach in conjunction with the lectures, [the lab] would be solid." These comments exactly reflect the type of scaffolding present in labatorials in an ideal situation, although labatorials do not require lab reports.

In summary, while peer interactions in traditional labs played a role in students' learning in the lab, the style of cooperation was typically not collaborative enough to scaffold students and promote deeper learning. The labs themselves did not emphasize learning concepts, though the theory explanation at the beginning of each lab did help frame students' thinking. Furthermore, one student in particular felt that the overall lab experience was not scaffolded enough, with the recipelike instructions of the lab not encouraging students to thoroughly think through the ideas of the lab. This, combined with the focus on error avoidance induced by the pressure due to grades and further encouraged by the recipe format, made the traditional labs generally less conducive to conceptual learning.

C. Analysis of student assessments

We now wish to examine the data collected throughout the study pertaining to students' conceptual understanding in order to answer RQ2b. Namely, in addition to examining the general features of labatorials and traditional labs, i.e., students' affective experience in each type of lab and the way each type of lab promotes or inhibits the development of conceptual understanding, we will now consider the key concepts to be learned in the course and analyze the performance and responses of each group as described in Sec. II D 3.

1. Establishing equivalence of the groups

As the first step in our analysis, we must establish the equivalence of the labatorial group and the traditional lab group. Namely, because we are not conducting pre-post testing, instead only comparing performance on parts of different assessments, the two groups must be equivalent at the start of the course in order for any differences detected later on to be meaningful. To do so, we compare the average total score for each group of the general pretest administered at the start of the course using a t test (see the "Conceptual Quiz on Classical Mechanics" dated May 7–9, 2019 found in the Supplemental Material [46]). The pretest

TABLE IV.   Descriptive statistics for labatorial and traditional lab pretest scores.

Lab group	Sample size	Mean score	Standard deviation
		Statistic	
Labatorial	30	2.130	1.676
Traditional	24	2.290	1.459

was graded out of six points, with only integer scores being possible due to the multiple choice nature of the questions. While this scale is evidently not continuous, we only use the t test as a starting point. The descriptive statistics for each group are shown in Table IV, and the score distribution of each group is shown in Fig. 3.

The t test yields a p value of 0.717, which means we cannot reject the null hypothesis that the means are equal. Additionally, a Mann-Whitney U test yields a p value of 0.703, which means that we cannot reject the null hypothesis that the distributions are the same. Considering that the means of the two groups are similar and the standard deviations are relatively large, it is reasonable to assume based on the large t test and Mann-Whitney U test p values that the two groups did not perform differently at a statistically significant level in the general pretest. As such, we proceed under the premise that the two groups are equivalent.

2. Preliminary statistical analysis

As previously described, the purpose of quantitatively comparing the performance of students in each group is to determine which questions are most worth qualitatively investigating. It is likely that not all questions will exhibit an interesting difference between the two groups, and so we wish focus on only those questions that can provide us with the most insight regarding the differences in students'
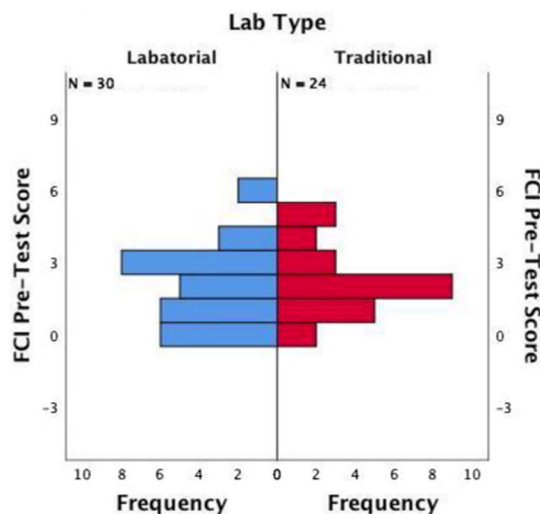

FIG. 3.   Labatorial and traditional lab pretest score distributions.

conceptual understanding. From this preliminary comparison of students' scores for each question, we can also hypothesize about which types of questions (e.g., conceptual, numerical, etc.) students perform better on in each group before proceeding to investigate this point more precisely with the qualitative analysis (and ultimately, the triangulation).

We present the test results only for the questions for which at least one of the tests yielded a significant result, i.e., a p value less than 0.05. Since t tests are being utilized extremely loosely in this portion of the analysis, we also consider questions for which the t test yielded a p value greater than but close to 0.05 since the small sample sizes (30 labatorial students and 24 traditional lab students) reduce the reliability of the results. For the question score distributions of the labatorial and traditional lab groups, see Supplemental Material [46]. The statistical test results and their tentative interpretations (i.e., our hypotheses) are summarized in Table V. Namely, labatorial students appear to perform better on conceptual questions, while traditional lab students appear to perform better on short calculation questions. Based on these results, we hypothesize that this holds in general, which we shall delve into by first examining the eight statistically analyzed questions qualitatively.

3. Qualitative analysis

In qualitatively analyzing students' post-test and final exam responses, we wish to gain a more nuanced understanding of the strong points and conceptual difficulties of labatorial and traditional lab students. This will allow us to begin formulating more general statements regarding the learning outcomes achieved by the students of each group and perform a meaningful comparison of their conceptual understanding. After establishing the student sample of interest for the analysis, an overview of the results will be shown and discussed.

In order to limit the scope of this analysis and avoid any possible biases, we concentrate on the responses to the eight selected questions of the 12 student interviewees as well as those of six additional students from each group. To that latter end, the additional students are selected with the aim of having a set of students from each group uniformly distributed across the associated course sections with equal gender representation that maximizes the diversity of final exam scores (and thus of possible levels of conceptual understanding) and student majors. We also ensure that at least one student from each labatorial working group in each section is included in the analysis in order to verify that the strong points or misconceptions of students are not isolated occurrences. The gender, class section, major, and final exam score of the selected students are summarized in Table VI.

To analyze question responses, we first separately compile all the responses of the appropriate students for

TABLE V. Summary of *t*-test and Mann-Whitney U test results for selected post-test and final exam and questions, with *p* values denoted by UEV indicating the usage of Welch's unequal variance *t* test rather than student's *t* test. The interpretations stated in the "stronger group" row are hypotheses for further qualitative exploration.

Results	Questions selected for analysis						
	Post Q1	Final Q5	Final Q9	Final Q10	Final Q14	Final Q15	Final Q17
t test	0.034 (UEV)	0.068	0.006 (UEV)	0.008	0.049 (UEV)	0.003	0.050 (UEV)
Mann-Whitney U	0.052	0.057	0.009	0.010	0.054	0.010	0.075
Stronger group	Labatorial	Traditional	Traditional	Labatorial	Labatorial	Traditional	Traditional
Question type	Concept	Short calculation	Short calculation	Concept	Concept	Short calculation	Short calculation

labatorials and traditional labs, which have already been formally graded with feedback. Then for each student, we briefly summarize and comment on the key conceptual difficulty (when appropriate) and the reasoning process (or lack thereof) exhibited in their response. From here, we separately summarize the key points that arose for the labatorial students and the traditional lab students (also counting the number of students who understood correctly in each) and then finally synthesize the results of the two groups.

For a detailed illustration of this approach as applied to the responses for question 10 of the final exam—a conceptual question pertaining to centripetal force and acceleration—see Supplemental Material [46]. In brief, the analysis indicates that the labatorial students understand the direction of the centripetal force and acceleration overall, with 8/12 students understanding centripetal force and 9/12 students understanding centripetal acceleration. Among the traditional lab students, similar conceptual errors were made overall, with 6/12 students understanding centripetal force and 4/6 understanding centripetal acceleration. While there are some common lingering misconceptions in both groups, errors occurred more frequently in the traditional group. The issue of drawing a tangential acceleration component in particular appeared to be more prominent among traditional lab students. Therefore, it seems likely that a labatorial worksheet has the ability to improve students' understanding of the concepts of centripetal force and acceleration in general compared to traditional lab students, although there is likely still room for improvement in the worksheet. This is consistent with the statistical test results of Table V.

A similar analysis was done for the other questions in Table V. For the full analyses, see Supplemental Material [46]. The analyses revealed several interesting points, which we will briefly discuss. There were several key concepts that labatorial students appeared to understand better than traditional lab students. For example, the responses to question 14 of the final exam and question 1.2 of the first post-test indicate that labatorial students may possess a stronger intuitive understanding of the meaning of the coefficient of restitution in collisions and the fluids concept of volume displacement, respectively. On the other hand, the traditional lab students perform better on questions 9, 15, and 17, which all involve one-step, formula-based calculations that students could refer to in their lab notebooks (or labatorial worksheets) during the exam.

TABLE VI. Metadata of labatorial and traditional lab students whose post-test and final exam responses are qualitatively analyzed.

Pseudonym	Gender	Section	Major	Final exam score
Extra 1	F	40	Religion	39
Jessica	F	44	Exercise Science	59
Catherine	F	40	Biology	65
Emma	F	42	Exercise Science	66
Quincy	M	40	Environmental Science	69
Derek	M	44	Behavioral Neuroscience	69
Extra 5	F	42	Environmental Geography	78
Extra 6	M	44	Independent Studies	78
Extra 2	M	40	Science and Technology	86
Extra 4	M	42	Exercise Science	89
Stacy	F	44	Biochemistry	91
Extra 3	M	42	Physics	92
Extra 6	F	43	Biology	41
Extra 1	M	41	Marketing	44
Extra 2	F	41	Psychology	58
Extra 4	F	41	Psychology	79
Extra 3	F	41	Chemistry	80
Oscar	M	41	Biology	82
Adrian	M	41	Exercise Science	82
Evelyn	F	43	Behavioral Neuroscience	85
Zion	M	43	Aerospace Engineering	87
Lauren	F	43	Behavioral Neuroscience	89
Amir	M	43	Chemistry	91
Extra 5	M	43	Software Engineering	91

Therefore, traditional lab students may be stronger at answering questions involving repetitive, template-based methods or tasks from the lab. However, based on question 15, many labatorial students appear to have a better understanding of the concept underlying these methods and exhibit more efforts at reasoning conceptually in their solutions. Lastly, the result from question 5 that traditional lab students are more proficient at data linearization than labatorial students is unexpected since the labatorial worksheets were designed to scaffold students' intuition on linearization. This may be because traditional students needed to do this as part of their reports in a more formulaic way, whereas labatorial students often just talked about the idea or sketched the graph, which may have affected how they absorbed the concept. It may also be necessary to refine the worksheets.

In summary, there is a tendency for labatorial students to possess a stronger intuitive understanding of the conceptual questions. On the other hand, traditional lab students tend to perform better on questions where short, formulaic calculations are involved. In addition, labatorial students may be attempting to think about the associated concepts more regardless of the question type, as suggested by the reasoning they exhibit in their solutions. However, this analysis does not consider all the post-test and final exam questions or even all the conceptual questions due to the heuristic we used to select the questions. As such, we will make a final conclusion regarding RQ2b on the learning outcomes of each group upon comprehensive triangulation with all the relevant data sources.

IV. TRIANGULATION

Recall that in Fig. 1, we were able to partition the aspects of our research questions into two categories: those that pertain to general characteristics of labatorials and traditional labs (RQ1 and RQ2a, colored in orange), and those that are directly tied to the content of the course curriculum (RQ2b, colored in blue). Furthermore, there are several data sources that can address each of those questions. However, thanks to the structure observed in our research questions, we were able to specify precisely which sources can be associated with which questions, giving rise to the two regions in the figure. Furthermore, within each of these sets of questions, one data source in particular provided the most information and thus could be considered the foundation of the data analysis addressing those questions: the student interviews for the general characteristics category, and the post-test and final exam written responses for the content-specific category.

We can now triangulate across the data sources in a way that takes into consideration the inherent structure of the data: the TA interviews, class observations, and TA surveys will be used to triangulate with the student interviews, and the class observations, TA surveys, student writing products, and analysis of students' question scores (similar to that described in Sec. II D 3) will be used to triangulate with the final exam and post-test responses. However, before we can perform triangulation, we must introduce the method used to analyze the sources of data that have not yet been considered.

A. Hierarchical summarization

We use the term "hierarchical summarization" to denote the approach taken in analyzing the class observations, TA surveys, and students' writing products. In this process, one needs to
(1) Code all the data.
(2) Summarize the coding results and implications for each student or TA (depending on the source of data) for each lab.
(3) Synthesize the results across all the students or TAs within each treatment group.
(4) Compare, contrast, and synthesize the results across each group (for each data type).
(5) Triangulate across all relevant data sources.
(6) Synthesize across all lab topics (if applicable).

The hierarchical nature of this framework is visualized and summarized in Fig. 4. For a demonstration of the first four steps of this procedure in the context of the lab on centripetal force (Lab 4), see Supplemental Material [46]. Because the final two steps of the hierarchical summarization procedure will be manifested differently for the student interviews and examination responses, they will be presented separately in the following two sections.

B. Triangulation of scaffolding and support themes

The student interview results will be triangulated with the TA interviews, classroom observations, and TA surveys. We begin by summarizing the themes that emerged from the TA interviews, which were analyzed using an inductive coding scheme as was done for the student interviews, for triangulating with the student interviews. These themes are summarized in Table VII and their relationships with the dimensions of the research question are illustrated using the concept map in Fig. 5. Any major similarities or discrepancies present in the classroom observations, which were analyzed using the aforementioned hierarchical summarization scheme, will be explicitly indicated. However, note that the TA surveys contained minimal information pertinent to the scaffolding aspect of the research, and the pertinent information was generally consistent with the other data. As such, they will not be explicitly discussed.

1. The student learning experience

In labatorials, the scaffolding provided by TAs had a strong impact on students' learning experience. Isaac noted that "it [took] more empathetic energy to understand what [students were] trying to do," which is consistent with students' comments about the TAs interactions with them
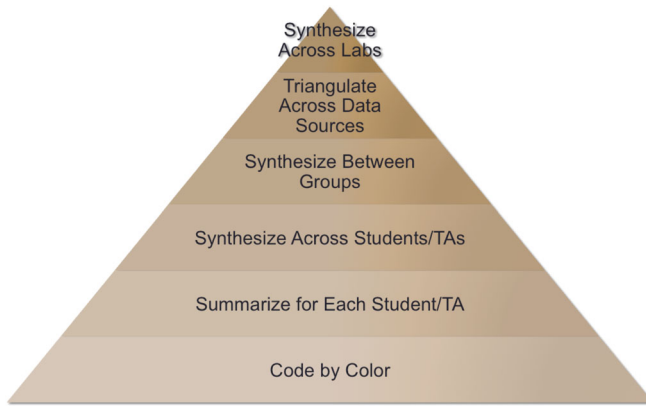
FIG. 4. Schematic of hierarchical summarization framework.

feeing very supportive and involved. Additionally, Liam stated that "if they [didn't] know how to solve [a problem], they [could] talk with [their peers]," which they typically did before going to the TA for help. Isaac and Justin also felt that students' interactions in labatorials were very collaborative overall. In particular, Isaac stated that "the majority saw that they really enjoyed working in teams" and that "they depend on each other for moving forward," which reinforces the affective notion of support as an aspect of the peer scaffolding mechanism in labatorials.

The course observations also corroborated these points overall. Students in general appeared to be comfortable sharing with each other what they did not understand, and due to this mutual support and the regular checking-in of the TA, students appeared relaxed for most of the labs. However, as with the TA interviews, nothing was noted regarding the support that students felt due to the grading scheme of the course, and so this theme cannot be triangulated. Furthermore, it was observed that some students exhibited untypical behavior such as keeping to themselves and dominating the pace of the group without checking in with their struggling teammates, and Stacy, a stronger student in the course, would sometimes experience frustration due to having to wait for the TA checkpoints in the worksheet. Therefore, while the authors believe the interpretation of students' learning experience as highly positive to be mostly representative of the class, there may additional student perspectives of labatorials yet to be explored.

On the other hand, the traditional lab students appeared to meaningfully communicate less frequently since "a lot of them had difficulty sharing ideas and explaining things to each other and asking questions," according to Justin. Isaac additionally suspected that an absence of group accountability could also be contributing to the lack of unity and the imbalances present in the teamwork in traditional labs:

They're not obliged to actually work well together [in traditional labs]. So you have more cases where there's this person who does everything, and the other ones are fine because there's no consequences to that.

This is consistent with the traditional lab students' comments expressing that the group work in traditional labs was often more akin to working independently than working collaboratively. As such, students would heavily rely on the TA for answering their questions. Isaac in particular noted this, saying that he had "way more interactions with the traditional [lab students] because they just [had] instructions" to follow. However, this is consistent with students' comments that the TAs were always very available and willing to help, which helped them feel at ease in the lab.

Indeed, it was observed that students would typically ask the TA questions as soon as something seemed amiss, typically relying on each other for more pragmatic tasks such as subdividing tasks and checking numerical results. Furthermore, these questions mostly involved clarifying instructions or how to use a formula. In conjunction with the observation that students were often engrossed in reading the manual, this may corroborate the student interview theme that they perceived the instructions as a source of support that, like the peer and TA support, could help them get the best grade possible.

In summary, triangulating across all the data sources pertinent to the students' learning experience (Step 5 of the hierarchical summarization for RQ1) indicated that the themes of support that emerged in the student interviews resonated across the data. The peer and TA interactions were positive in nature in both labatorials in traditional labs, the nature of the associated support differed between the two approaches, with those in labatorials emphasizing collaboration and camaraderie and those in traditional labs

TABLE VII. Summary of themes from TA interviews.

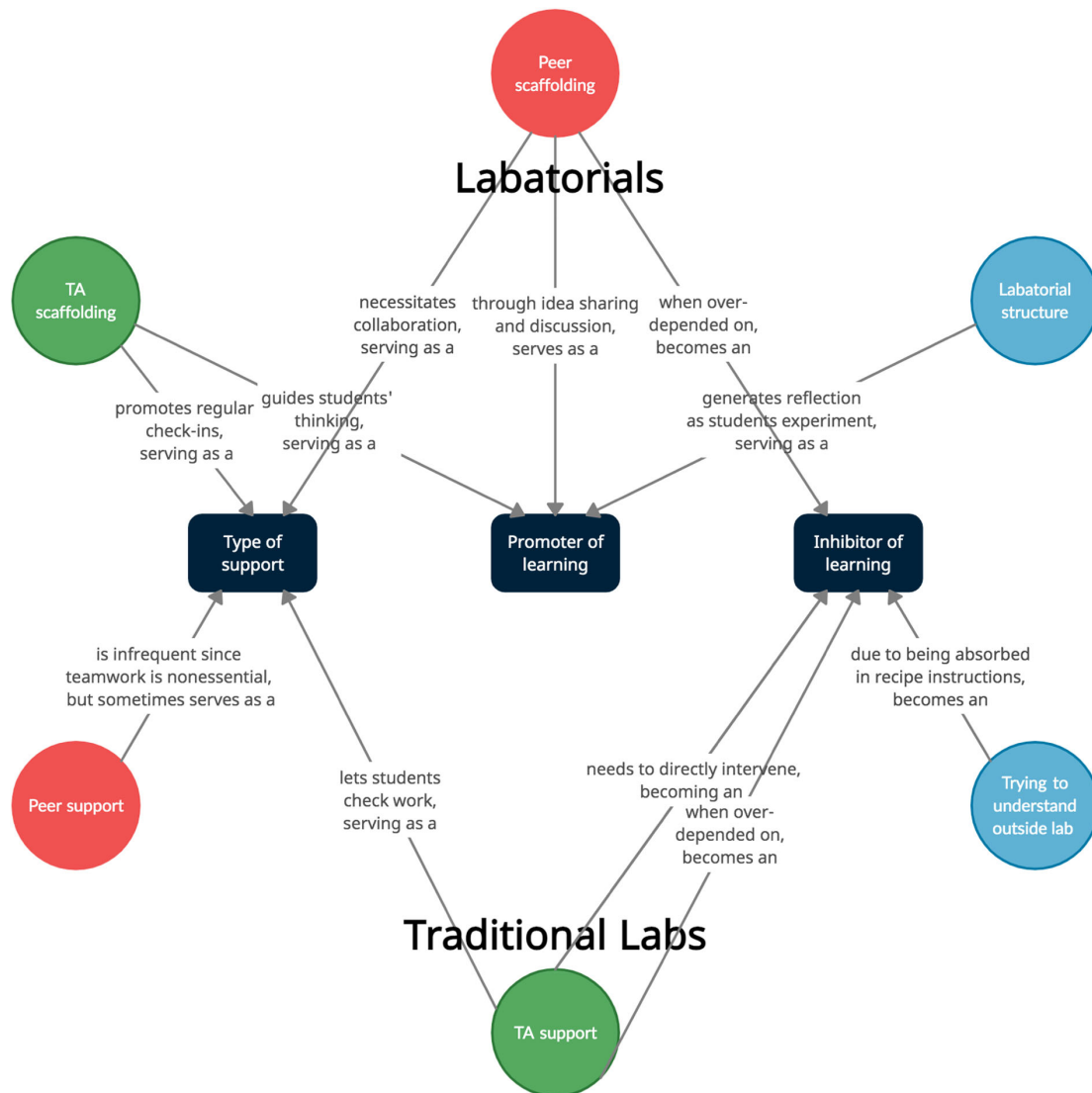| Lab type	Dimension of research questions		
	Support types	Learning promoters	Learning inhibitors
Labatorial	Peer scaffolding TA scaffolding	Peer instruction TA scaffolding Labatorial structure	Peer overdependence
Traditional	Peer support TA support		TA over-dependence Frequent, direct intervention Trying to understand after lab is already done

FIG. 5. A concept map summarizing the key relationships between the TA interview themes and the dimensions of the research questions. Red, green, and blue nodes indicate overarching themes broadly related to peer interactions, TA interactions, and lab structure, respectively. Each arrow corresponds one of the specific themes listed in Table VII.

emphasizing pragmatism and individualism despite being in a group. This difference appeared to arise due to the different behaviors encouraged by the format of each type of lab: sharing ideas and discussing in labatorials, and following instructions in traditional labs.

2. Conceptual learning in the lab

Just as peer interactions were important for labatorial students on an affective level, labatorials also played a large role in students' conceptual learning. A sentiment shared by all the TAs, Isaac indicated that peer instruction frequently occurred by saying that the way teammates in labatorials depended on each other "[pushed] them to also share the information that they [understood]," and that "by sharing, by putting what they know into words, they

understand better." As also observed by the author (F. L.), Justin witnessed many light-bulb moments following involved student discussion, leading him to express that the teamwork—wherein students would "figure [their confusions] out between themselves"—was the strongest aspect of labatorials. Nevertheless, the author also observed that some labatorial students did at times overdepend on their peers, "not really understanding everything and […] just copying from their teammates, just being carried by their team," as stated by Justin. This is consistent with the struggles of students like Emma as discussed in the student interview analysis.

As was also evident in the student interviews and course observations, all three TAs indicated that they took a scaffolding approach in the lab. Like Liam, all the TAs indicated that they always "[tried] to guide them about how

to solve [a] problem." Consistent with the TA interviews, this scaffolding was observed to sometimes be in the form of giving a hint or encouraging them to further explore an idea they had. As stated by Isaac, this ensured that students "at least […] went through the effort [of] trying to figure it out, even [if] they were wrong," which appeared to promote the regular group discussion and engagement with the concepts that was observed by the author and greatly helped students learn. When listening to those discussions and the subsequent interactions with the TA, the author noticed that students' questions for the TA were more subtle, also noted by Justin:

> *The quality of their questions was better in labatorials because before they [asked] questions, they [had] already done some discussion between themselves. So they at least [came] up with one hypothesis or whatever and then [asked] me questions. But in the traditional labs, most of the time the questions they [asked were] very trivial and not always well thought through.*

However, both the TAs and the author noted that there were also cases of students having trouble setting up the experiments and not performing the experimental procedures completely rigorously. This suggests a possible disadvantage of labatorials focusing on concepts over experimental skills. However, this may also simply indicate that changes to the worksheets are required. Furthermore, Liam had the opposite impression, and so this issue may only apply to some students rather than be a side effect of labatorials. Additional research is necessary to address this point.

In contrast, as Justin's prior comment indicates, the traditional lab students were not as engaged in their learning. As a result, the TA served not to help students learn, but to check their results and their procedures, as with their peers. This overdependence on the TA caused them to not think about the concepts deeply, which could have inhibited their learning. Isaac indicates that this was due in part to the recipelike format of the labs:

> *With [traditional lab students], they have been given some instructions, they just follow mechanically, they do it, and then they ask if something is wrong. But they're not really understanding. I mean, not all of them, but most of them I would say they just approach it in this mechanical way, doing this and this and that and that, and at the end you get a couple of numbers.*

Nevertheless, there were few instances of students discussing concepts with each other and asking the TA conceptual questions, consistent with the author's observations. Isaac additionally commented that he felt that he was "intervening all the time" in traditional labs "because [they] have to manipulate the instruments a lot more, and if [he doesn't] intervene, they never finish on time." As such,

this type of frequent, direct intervention may also be deterring students from deeply thinking through the lab. Both of these behaviors are consistent with the theme of focusing on error avoidance that emerged from the student interviews and act as an inhibitor to learning.

The TAs also indicated that traditional lab students were not focused on their learning in the lab, but were instead absorbed in the recipelike lab instructions and focused on performing all the steps as correctly as possible, as also observed by the author. As such, they may only have begun thinking about the concepts by the time they wrote their reports and then grappled with the concepts while cramming for the final exam, with Isaac stating that "weren't understanding while [they] were doing it." Because students waited until after the lab to attempt understanding the concepts and the lab as a whole, they will likely not have discovered any misconceptions they may have had about the concepts, rendering their learning less effective. This observation is consistent with labatorial students' comments on the matter with regard to their prior traditional labs.

In summary, triangulating across all the data sources pertinent to the content-independent aspects of students' conceptual learning in the lab (Step 5 of the hierarchical summarization for RQ2a) indicated that labatorials have several mechanisms of scaffolding put in place that mutually enhance each other; students were more engaged with each other and the material, allowing them to resolve conceptual difficulties and to have more fruitful interactions with the TA who was himself encouraging students to collaborate. This was all supported by the structure of the labatorials by having the checkpoints and the gradual building-up of concepts integrated into the worksheets. Contrariwise, the TAs did not refer to any particular promoters of learning in traditional labs, instead identifying inhibitors strongly related to those indicated by students. As such, all three TAs felt that labatorial students learned the concepts more deeply than the traditional lab students, with Isaac in particular stating that "[he] would actually put money that they [understood] better in labatorials because of the [lab report] conclusions that [he] read from the traditional lab [students]." However, the impact of labatorials on students' experimental skills needs to be further explored.

C. Triangulation of student learning outcomes

It is not immediately clear from each of the qualitative data sources pertinent to examining students' learning outcomes (i.e., class observations, TA surveys, and student writing products) how they can best be triangulated with the final exam and post-test responses. To this end, we outline an additional triangulation heuristic that can be used to synthesize all the analyses (both qualitative and quantitative) and begin extracting general results about students' learning of the concepts, addressing steps 5 and 6 of our

hierarchical summarization framework. Upon completing these steps, we will be able to make connections between the presence (or absence) of scaffolding mechanisms in labs and their impact on students' learning outcomes.

1. Developing a triangulation heuristic

As there are several important dimensions to consider in order to triangulate students' learning outcomes across different data sources (lab type, learning outcome, data source, and concept), we structure the results of steps 1–4 of our hierarchical summarization procedure as a multi-layered cross tabulation between these dimensions. In particular, in order to visualize the prominence of the key conceptual gains and difficulties experienced by students and any patterns therein, we count the number of occurrences (denoted by stars) of gain or difficulty for each concept exhibited in each data source and enter this count into the appropriate region of the cross tabulation, denoted labatorial gain, labatorial difficulty, traditional gain, or traditional difficulty. As can be seen in Fig. 6, this effectively results in a quadrant system (whose dimensions are the lab type and the overall learning outcome), dividing the figure into four smaller cross tabulations recording the occurrences (whose dimensions are the data type and the identified concept).

In order to take into account the whole class's performance on a topic in addition to counting the occurrences of conceptual gain or difficulty for the selected final exam or post-test questions in Table V, we include a column containing the post-test (P) and final exam (F) scores (XX%) for each question (whose number is appended to P or F) related to the given concept, not only those that were analyzed qualitatively. (Since there is only one post-test for each lab, the post-test number is the same as the lab number. As such, P2(94) in Fig. 6, which is for Lab 4, should be interpreted as a score of 94% on question 2 of Post-Test 4.) This column functions under its own heuristic; if the scores are significantly different between the two groups (based on the analysis summarized in Table V), then the scores are entered into diagonally opposite quadrants of the appropriate groups (i.e., labatorial gain and traditional difficulty or labatorial difficulty and traditional gain). On the other hand, if there is no statistically significant difference, then because all the final exam averages range from approximately 40% to 100%, we choose 70% as a reasonable (albeit quasiarbitrary) threshold for placing each group's performance into a conceptual gain or conceptual difficulty quadrant.

In order to now derive meaningful information from this figure, we must discuss how such a representation of data should be interpreted. First and foremost, since we are trying to make comparisons between labatorials and traditional labs, entries in Fig. 6 will only be meaningful if they allow us to conclude that the two student groups are the same (in terms of their average understanding) or different. In order for one to draw the former conclusion, there must be occurrences uniformly distributed across all quadrants of the figure. This is because while both groups only having occurrences in their respective "gain" or "difficulty" quadrants may imply that they both understand (or do not understand) the concept, the lack of occurrences in the opposite quadrants may simply be coincidental (i.e., faults in understanding may simply not have been detected in our data). Therefore, in order for one to draw the latter conclusion from the figure, there must be an overdensity of occurrences in diagonally opposite quadrants, with the most conclusive case being that the occurrences of a concept are strictly in diagonally opposite quadrants. For a detailed discussion of the considerations taken into account to ensure the reliability of this triangulation scheme and any heuristics utilized therein, see Supplemental Material [46].

With these considerations in mind, we refine the triangulation cross-tabulation shown in Fig. 6 so as to only include concepts from which some conclusion can be drawn, resulting in the reduced quadrant representation shown in Fig. 7. This shows occurrences spread across the four quadrants, with the highest density in the labatorial-gain and traditional-difficulty quadrants. Therefore, we can conclude with reasonable confidence that the labatorial students' understanding of centripetal force is stronger, a conclusion supported by the closer examinations of the tabulated data. For similar reasons, we can make the same conclusion for the concept of centripetal acceleration. Contrariwise, Fig. 7 suggests that traditional lab students better understood the concepts of graph linearization and line of best fit. This analysis, which is a part of step 5 of our hierarchical summarization procedure, can be conducted similarly for the other lab topics.

2. Notable student learning outcomes

Upon triangulating between the various data sources for each lab as just described, we complete step 5 of the hierarchical summarization and compile the results in Table VIII according to the lab number and which group exhibited greater mastery of the concept or skill.

When examining these results further, a pattern emerges in terms of the types of concepts that students of each group tend to perform well across all labs, allowing us to achieve step 6 of the hierarchical summarization procedure: the labatorial group exhibits mastery of core lab concepts, while the traditional lab group exhibits mastery of standardized procedures and memorization-based calculations. These results both make sense considering the focus of each type of lab; labatorials are designed to reinforce concepts, while traditional labs involve recipelike work.

		Labatorial					Traditional				
		Obs.	TA Survey	Work-sheet	Final/Post	Scores	Obs.	TA Survey	Report	Final/Post	Scores
Conceptual Gain	Relationship between Fc and v, r		*			P2(94)		*			P2(90)
	Fc direction			*	****	P1(83) F10(80)					P1(72)
	a direction			******	****	F10(80)					
	v direction			****							
	T = Fc + Fg			****							
	Inertia			*							
	Linearization				**				**	****	F5(62)
	Line of best fit				***					****	F17(82)
	Sources of error							**			
Conceptual Difficulty	Inertia		*	**							
	T = Fc + Fg	*	*	**							
	F direction	*		*	*					*	F10(61)
	a direction			*	*					**	F10(61)
	Line of best fit				**	F17(59)	*	*		**	
	Sources of error								***		
	Focused on formula verification								■		
	Linearization				****	F5(41)		*	*	**	

FIG. 6. Raw compilation of occurrences of conceptual gain and difficulty for labatorial and traditional lab students in lab 4 across data sources. The number of stars in a cell represents the number of times signs of conceptual gain or loss were identified in the corresponding data source (column). The different colors are used as a means of quickly identifying occurrences of corresponding concepts (rows) in the different quadrants. For the "scores" column, the letter P/F is used to denote whether the score (XX%) is from a post-test or final exam.

Although all students had access to resources during the final exam (i.e., the worksheets for the labatorial students and the lab notebooks for the traditional lab students) with all the same essential information and equations, the labatorial students did not seem to be as proficient at the questions about linearization or plotting a line of best fit, for example. In addition to the possible cause of labatorial students' minor graphing errors during the labs not always getting caught by the TA, there is also the possibility that due to the nature of the questions in the worksheet, they would try and explain more than needed or overthink somehow and thus make a mistake. This also occurred quite often in the one-step, formula-based exam questions, and so the underlying cause of the discrepancy may be the same. However, it does seem reasonable that despite this, both groups perform equally well on the longer calculation questions; since they cannot be solved by pure memorization (e.g., questions 3 and 8 of the final exam) and both groups performed similar amounts of such calculations during the lab sessions, we would not expect a difference.

While we would have expected labatorial students to perform better than traditional lab students on all of the conceptual questions, there could be a multitude of reasons pertaining to the design of the course or the questions themselves that could have rendered this implementation of the course suboptimal. The results nevertheless serve as evidence that the extensive scaffolding mechanisms present in labatorials are indeed improving students' conceptual learning in the lab. Furthermore, they suggest a general trend for the types of questions that each type of lab teaches students to solve, which is a valuable result since it may hold generally for the two lab types regardless of the specific course content. However, additional research would be required to confirm this.

V. CONCLUSIONS

In this study, we explored how the experience of learning differs between labatorials and traditional labs. In particular, we were interested in the ways in which students' affective learning experience could differ between the two

		Labatorial					Traditional				
		Obs.	TA Survey	Work-sheet	Final/Post	Scores	Obs.	TA Survey	Report	Final/Post	Scores
Conceptual Gain	*F direction*			*	****	P1(83) F10(80)				***	P1(72)
	a direction			******	****	F10(80)				**	
	Linearization				**			**		****	F5(62)
	Line of best fit				***					****	F17(82)
Conceptual Difficulty	*F direction*	*		*	**					***	F10(61)
	a direction			*	*					****	F10(61)
	Linearization				****	F5(41)	*		*	*	
	Line of best fit				**	F17(59)	*	*		**	

FIG. 7. Refined compilation of occurrences of conceptual gain and difficulty for labatorial and traditional lab students in lab 4. The notation used is the same as that in Fig. 6. The colors again are used as a means of quickly identifying occurrences of corresponding concepts (rows) in the different quadrants, but the specific colors chosen are arbitrary, and so there is no connection to the colors used in Fig. 6.

types of labs as well as the ways in which they each promoted the development of conceptual understanding. It was found that students' learning experience as well as the quality of their conceptual learning was generally better in labatorials than in traditional labs by virtue of three key forms of scaffolding present in labatorials that were affectively perceived by students as forms of support: peer scaffolding, instructor scaffolding, and scaffolding by the labatorial worksheet. These contributed to the clear differentiation of the learning outcomes of labatorial and traditional lab students: labatorial students exhibited greater mastery of concepts, whereas traditional lab students exhibited greater mastery of short, template-based calculations and procedures.

A. Summary of key results

In establishing the three overarching themes of peer scaffolding, instructor scaffolding, and labatorial worksheet scaffolding from the labatorial student interviews, certain aspects of the lab course were found to impact students' learning experience and conceptual learning, the structure of which was illustrated in Table III and Fig. 2 (similarly for traditional labs). The highly collaborative nature of labatorials allowed peers to support each other in the lab. Students felt a sense of camaraderie in working with their peers, feeling like it was always a team effort toward understanding. This helped them feel more comfortable sharing doubts about their knowledge and have a more positive overall experience. This sentiment was reinforced by the interactions with the TA, which—through his regular guidance at checkpoints and general involvement with the students—allowed students to feel more confident in working through the worksheet since they knew he was there to scaffold their understanding. Moreover, because of the low-stakes grading of labatorials, students were able to not worry about making mistakes, instead focusing on the

learning experience and feeling more fulfilled upon completing the lab. This, in conjunction with the students' perceived relevance of the content beyond the lecture course, lead many students to express an improved outlook on physics in general, indicative of a highly positive affective transformation.

By virtue of this peer, TA, and graded-related support, labatorial students were also able to get more engaged with the concepts and the learning process in the lab, regularly getting involved in discussions with their peers and making an effort to understand the concepts. While there was the possibility of weaker students sometimes depending on the stronger students for progressing, the peer instruction was valuable to students not only for figuring out problems, but also—by virtue of the structure of the labatorial worksheets —for effectively dealing with the cognitive dissonance involved in overcoming misconceptions. The worksheet elicited their prior ideas, the experiments performed challenged those preconceptions, and then peers worked together to resolve their conceptual inconsistencies. The strategically located checkpoints of the labatorials also ensured that the TA could verify their state of understanding and then guide their thinking at key parts of the worksheet. By virtue of these various forms of scaffolding, labatorial students exhibited a deeper conceptual understanding of the core lab concepts than the traditional lab students, typically performing better on the conceptual questions on post-tests and the final exam. Interviewees also all expressed a strong general preference for labatorials over the traditional approach.

In traditional labs, while there were mechanisms of support in place that helped improve students' learning experience, scaffolding was less prominent than in labatorials in all respects. The peer support in traditional labs typically involved verifying for correctness of each other's results and did not result in true collaboration. The TA

TABLE VIII.   Concepts best learned by labatorial and traditional students.

Lab topic	Lab type		
	Labatorials	Equivalent	Traditional labs
1. Density of solids or liquids	Volume displacement	Density equation Sources of error	
2. Vector addition of forces		Graphical methods Component method Equilibrant force	
3. Springs	Spring mass fraction	Period equation Hooke's law	Mass fraction graphical calculation
4. Circular motion	Force direction Acceleration direction	Centripetal force equation usage	Graph linearization Line of best fit
5. Coefficient of restitution	Interpretation of coefficient of restitution KE loss in collisions	Free-fall math	KE loss math
6. Pendulum motion	Period intuition	Period equation	Graph linearization

played a similar role for students, being available whenever they had procedural questions about the experiment. The TA would also occasionally check in on students to see if they were doing things correctly, and so their role was largely managerial as opposed to one of guidance. These types of relatively shallow interactions, although reassuring for students, were incurred by a pressure to follow the instructions correctly in order to get a good grade. As such, for some students the recipelike instructions themselves served as a source of support. While most students did not express having a particularly negative lab experience by virtue of these forms of support, there were no unprompted statements of satisfaction, and the overall experience of working in the lab was strongly influenced by the recipelike lab format.

This focus on the lab instructions and error avoidance also had adverse effects on their learning. Rather than thinking about the concepts behind the experiments, traditional lab students were largely engrossed in their instructions, typically not engaging in any meaningful discussion with their peers or the TA and only trying to understand the experiment once they began writing their lab reports. Many students felt that the introductory theory explanation at the beginning of the lab helped them to begin thinking about the concepts and that seeing the concepts in a more tangible way via the experiments helped reinforce their prior understanding. However, misconceptions were less likely to be elicited and confronted in this framework, and so cognitive dissonance would not have been resolved. This absence of the conditions for the development of conceptual understanding effectively encouraged students to simply follow instructions and proceed through the lab without thinking about what they were doing. The effect of this repetitive, procedural nature of traditional labs on students' learning was also apparent in the final exam of the course; the traditional lab students typically performed

better on questions that involved short, template-based calculations or formulaic procedures.

Upon considering these results as a whole, there are clear dichotomies between labatorials and traditional labs that emerge regarding the forms of support in the lab, the pedagogical approaches taken, and the resultant impact of these on students' conceptual learning, which are summarized in Table IX. These are a result of the stark contrast between the emphasis on concept verification, detailed step-by-step procedures, and a largely supervisory instructor role in traditional labs and the three forms scaffolding at the forefront of labatorials:

(1) **Peer scaffolding:** At the forefront is the collaborative learning aspect of labatorials, wherein students actively engage in peer-instruction. Because students often come from different backgrounds and each possess their own strengths and weaknesses, peer instruction is a natural process in such an instructional setting, with students being scaffolded by their more knowledgeable peers and helping each other proceed through the worksheet.

(2) **Instructor scaffolding:** This role of the instructor is critical since, being an expert in the subject matter of the course, they allow the students to examine their understanding when encountering some difficulty or arriving at a checkpoint. Additionally, as an educator, they are in the best position to guide students' thinking in the right direction and to help arrange their learning experience so as to optimize their learning.

(3) **Structural scaffolding:** As with the tutorials that inspired them, labatorial worksheets are structured such that they elicit students' current understanding of a concept to the forefront via the prediction questions and allow them to acknowledge any possible inconsistencies. The worksheets then allow students to confront their misconceptions

TABLE IX. Dichotomies between labatorials and traditional labs identified from student interviews and assessments triangulated with secondary data sources.

Dimension	Lab type	
	Labatorials	Traditional labs
Lab focus	Conceptual	Experimental
Student focus	Learning	Error avoidance
Teamwork style	Collaborative	Independent
Accountability	Group	Individual
TA Involvement	Guidance or collaborative	Managerial
Lab structure	Scaffolding	Recipe
In-Lab	Understanding while doing	Doing without understanding
Learning outcomes	Conceptual understanding	Formulaic procedures

via experiments and ultimately resolve their cognitive dissonance and construct a full understanding by virtue of the progressive nature of the questions and the other forms of scaffolding in the lab.

B. Limitations and recommendations for future research

One major limitation is due to sample size, in terms of both the number of student interviewees and the total number of students in the course. Although the student interviews constitute the core of much of the qualitative portion of this analysis, the results derived are limited by the number of perspectives upon which they are based; differently stated, they are limited by the diversity of the interviewees. This is the main weakness of interviewing participants on a volunteer basis since one might expect certain types of students to be more willing to volunteer than others. Similarly, because the qualitative analysis of the post-test and final exam responses is based only on a subset of the questions—of which only the responses of 24 out of 54 students were examined—there may be interesting results that were missed in the unchecked responses that could have influenced our conclusions. The quantitative analysis is also limited; because the class size is small, the power of the statistical tests used is reduced. Furthermore, t tests may not necessarily give information as to the practical significance of quantitative results, although this portion of the analysis was secondary and so does not hinder our conclusions.

There are also several avenues for furthering the results of this work and for exploration that may be of interest to the greater research community. While not discussed in this work, labatorial students appeared to exhibit more thorough reasoning in their exam question responses as well as attempt to explain their thought process more often than traditional lab students. One traditional lab student's response in particular also suggested that traditional lab students may be less willing to think through new situations, perhaps due to a template-seeking mindset incurred by their lab experiences. These interesting trends

in students' thinking and approaches to problem solving could be more deeply explored in a future work, perhaps designing interview questions that explicitly elicit students' reasoning abilities as well as their views toward problem solving in physics for each type of lab.

Another interesting possibility includes examining labatorials and traditional labs in the context of other physics courses in order to see if the general conceptual learning trends posited in this work also hold for other topics. There may also be several topic-specific nuances that could affect the way one considers the design of labatorials. While labatorials show much promise for any sub-discipline (and possibly any discipline), all existing work on labatorials has focused on the introductory level. Therefore, the range of applicability of labatorials could be investigated by implementing them in mid- and upper-undergraduate physics lab courses, for example.

By virtue of this study's implementation of labatorials separate from a lecture course, we were able to explicitly show for the first time that students learn concepts more meaningfully and deeply in labatorials than in traditional labs, all the while genuinely enjoying the experience. By explicit comparison of two different groups within a same course, the conditions for improving conceptual understanding in labatorials were identified in terms of the constructivist notion of scaffolding while simultaneously confirming the absence of these conditions in traditional labs. In the future, we hope that this comparative study will serve as a source of inspiration for physics educators who wish to make simple but powerful changes in their classrooms as well as a foundation for further exploration in conceptually driven physics lab pedagogy, which has the potential not only to deepen our understanding of students' learning in a lab setting, but also to impart to students a superior education in an aspect of physics fundamental to the nature of the discipline.

ACKNOWLEDGMENTS

[1] L. Kirkup, S. Johnson, E. Hazel, R. Cheary, D. Green, P. Swift, and W. Holliday, Designing a new physics laboratory programme for first-year engineering students, Phys. Educ. **33**, 258 (1998).

[2] M. Hanif, P. H. Sneddon, F. M. Al-Ahmadi, and N. Reid, The perceptions, views and opinions of university students about physics learning during undergraduate laboratory work, Eur. J. Phys. **30**, 85 (2008).

[3] M. D. Sharma, A. Mendez, I. M. Sefton, and J. Khachan, Student evaluation of research projects in a first-year physics laboratory, Eur. J. Phys. **35**, 025004 (2014).

[4] P. Aceituno, J. Hernández-Aceituno, and A. Hernández-Cabrera, Simulation of general physics laboratory exercise, in J. Phys. Conf. Ser. **574**, 012068 (2015).

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

[6] W. M. Roth, Experimenting in a constructivist high school physics laboratory, J. Res. Sci. Teach. **31**, 197 (1994).

[7] A. Hofstein and V. N. Lunetta, The laboratory in science education: Foundations for the twenty-first century, Sci. Educ. **88**, 28 (2004).

[8] D. Ahrensmeier, A practical application of Physics Education Research-informed teaching interventions in a first-year physics service course, J. Tech. Educ. **1**, 166 (2013).

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

[10] J. Lochhead and J. Collura, A cure for cookbook laboratories, Phys. Teach. **19**, 46 (1981).

[11] R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the evaluation of active learning laboratory and lecture curricula, Am. J. Phys. **66**, 338 (1998).

[12] N. Holmes, J. Olsen, J. L. Thomas, and C. E. Wieman, Value added or misattributed? a multi-institution study on the educational benefit of labs for reinforcing physics content, Phys. Rev. Phys. Educ. Res. **13**, 010129 (2017).

[13] C. Deacon and A. Hajek, Student perceptions of the value of physics laboratories, Int. J. Sci. Educ. **33**, 943 (2011).

[14] M. M. Tlowana, Student perceptions of the introductory physics laboratory: An exploratory study, Ph. D. thesis, University of Cape Town, 2017.

[15] D. R. Sokoloff, P. W. Laws, and R. K. Thornton, Realtime physics: Active learning labs transforming the introductory laboratory, Eur. J. Phys. **28**, S83 (2007).

[16] D. Ahrensmeier, J. Donev, R. Hicks, A. Louro, L. Sangalli, R. Stafford, and R. Thompson, Labatorials at the University of Calgary: In pursuit of effective small group instruction within large registration physics service courses, Phys. Canada **65**, 214 (2009).

[17] L. C. McDermott and P. S. Schaffer, *Tutorials in Introductory Physics* (Pearson, London, United Kingdom, 2002).

[18] E. Mazur and R. C. Hilborn, Peer instruction: A user's manual, Phys. Today **50**, No. 4, 68 (1997).

[19] C. H. Crouch and E. Mazur, Peer Instruction: Ten years of experience and results, Am. J. Phys. **69**, 970 (2001).

[20] E. Etkina, A. Karelina, M. Ruibal-Villasenor, D. Rosengrant, R. Jordan, and C. E. Hmelo-Silver, Design and reflection help students develop scientific abilities: Learning in introductory physics laboratories, J. Learn. Sci. **19**, 54 (2010).

[21] J. S. Bruner, *The Process of Education* (Harvard University Press, Boston, MA, 2009).

[22] J. S. Bruner, *The Culture of Education* (Harvard University Press, Boston, MA, 1996).

[23] J. S. Bruner *et al.*, *Toward a Theory of Instruction* (Harvard University Press, Boston, MA, 1966), Vol. 59.

[24] L. S. Vygotsky, in *Mind in Society: The Development of Higher Psychological Processes*, edited by M. Cole, V. John-Steiner, S. Scribner, and E. Souberman (Harvard University Press, Cambridge, MA, 1980).

[25] L. Festinger, *A Theory of Cognitive Dissonance* (Stanford University Press, Stanford, CA, 1962), Vol. 2.

[26] J. Dewey, *How We Think* (Courier Corporation, Mineola, NY, 1997).

[27] W.-M. Roth and A. Jornet, Toward a theory of experience, Sci. Educ. **98**, 106 (2014).

[28] J. I. Rotgans and H. G. Schmidt, Cognitive engagement in the problem-based learning classroom, Adv. Health Sci. Educ. **16**, 465 (2011).

[29] T. Bell, D. Urhahne, S. Schanze, and R. Ploetzner, Collaborative inquiry learning: Models, tools, and challenges, Int. J. Sci. Educ. **32**, 349 (2010).

[30] M. Pedaste, M. Mäeots, L. A. Siiman, T. De Jong, S. A. Van Riesen, E. T. Kamp, C. C. Manoli, Z. C. Zacharia, and

E. Tsourlidaki, Phases of inquiry-based learning: Definitions and the inquiry cycle, Educ. Res. Rev. **14**, 47 (2015).

[31] C. C. Kuhlthau, L. K. Maniotes, and A. K. Caspari, *Guided inquiry: Learning in the 21st Century: Learning in the 21st Century* (ABC-CLIO, Santa Barbara, CA, 2015).

[32] J. Allen, L. Barker, and J. Ramsden, Guided inquiry laboratory, J. Chem. Educ. **63**, 533 (1986).

[33] D. Ahrensmeier, R. I. Thompson, W. J. Wilson, and M. Potter, Labatorials—A new approach to teaching electricity and magnetism to students in engineering, in *Proceedings of the 2012 IEEE International Symposium on Antennas and Propagation* (IEEE, Bellingham, WA, 2012), pp. 1–2.

[34] C. S. Kalman, Enhancing students' conceptual understanding by engaging science text with reflective writing as a hermeneutical circle, Science & Education **20**, 159 (2011).

[35] C. S. Kalman, M. Sobhanzadeh, R. Thompson, A. Ibrahim, and X. Wang, Combination of interventions can change students? epistemological beliefs, Phys. Rev. ST Phys. Educ. Res. **11**, 020136 (2015).

[36] M. Sobhanzadeh, C. S. Kalman, and R. Thompson, Labatorials in introductory physics courses, Eur. J. Phys. **38**, 065702 (2017).

[37] C. S. Kalman, J. El-Helou, and M. J. Lattery, Improving high school students' understanding of the concept of force and Newton's laws through the combination of labatorials and reflective writing, in *Proceedings of the Annual NARST Conference March 31-April 3, 2019* (NARST, 2019), https://peer.asee.org/34314.

[38] J. B. Biggs and K. F. Collis, *Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of the Observed Learning Outcome)* (Academic Press, New York, NY, 2014), p. 10.

[39] R. M. Felder and R. Brent, Cooperative learning in technical courses: Procedures, pitfalls, and payoffs, ERIC Document Reproduction Service ED 377038, 1–18 (1994).

[40] C. Lee and J.-L. Farh, Joint effects of group efficacy and gender diversity on group cohesion and performance, Applied Psychol. **53**, 136 (2004).

[41] R. W.-Y. Cheng, S.-F. Lam, and J. C. Chan, When high achievers and low achievers work in the same group: The roles of group heterogeneity and processes in project-based learning, Br. J. Educ. Psychol. **78**, 205 (2008).

[42] W. Miller and H. J. Otto, Analysis of experimental studies in homogeneous grouping, J. Educ. Res. **21**, 95 (1930).

[43] J. McGaughy, Homogeneous grouping of pupils, Childhood Educ. **6**, 291 (1930).

[44] P. C. Blumenfeld, R. W. Marx, E. Soloway, and J. Krajcik, Learning with peers: From small group cooperation to collaborative communities, Educ. Res. **25**, 37 (1996).

[45] D. Esposito, Homogeneous and heterogeneous ability grouping: Principal findings and implications for evaluating and designing more effective educational environments, Rev. Educ. Res. **43**, 163 (1973).

[46] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.17.010131 for additional analysis and discussion, sample labatorial on Centripetal Force, interview questions, sample observation data, other qualitative data collection materials, and secondary quantitative data.

[47] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, Phys. Teach. **30**, 141 (1992).

[48] J. Corbin and A. Strauss, *Basics of qualitative research: Techniques and procedures for developing grounded theory* (Sage Publications, Thousand Oaks, CA, 2015).

[49] G. D. Ruxton, The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test, Behav. Ecol. **17**, 688 (2006).

[50] P. E. McKnight and J. Najab, Mann-Whitney U Test, in *The Corsini Encyclopedia of Psychology* (2010), pp. 1–1.