Introductory physics labs: WE CAN DO BETTER

Research reveals that labs are more effective when their goal is to teach experimental practices rather than to reinforce classroom instruction.

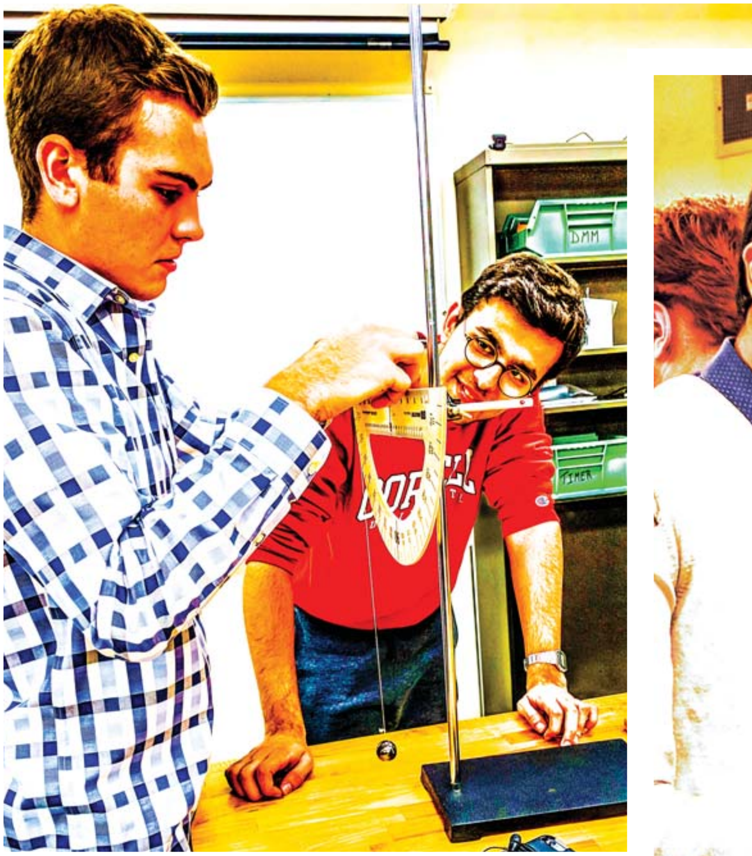
Natasha G. Holmes and Carl E. Wieman

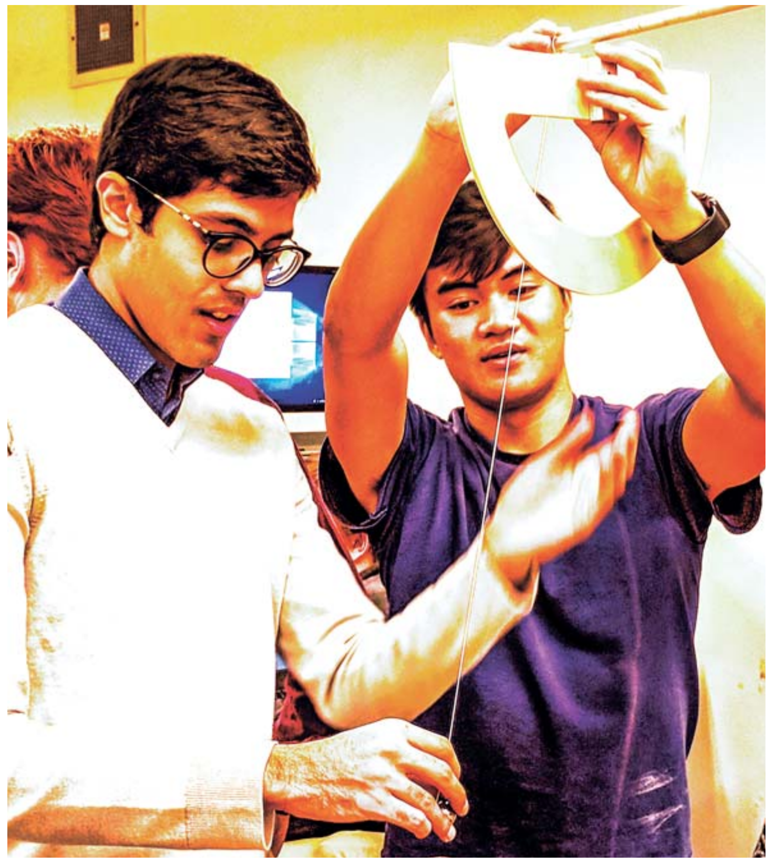
Lab instruction has historically been a cornerstone of physics education (see the article by Valerie Otero and David Meltzer, PHYSICS TODAY, May 2017, page 50), but the large amounts of money, space, and instructor time that labs require must be constantly justified. Few physicists would contest the importance of labs in a physics curriculum. But if you ask physicists what labs are for, their answers will be vastly varied. Goals range over reinforcing content, learning about measurement and uncertainty, practicing communication skills, developing teamwork skills, and, more broadly, learning that physics is an experimental science. Some labs try unrealistically to hit all those targets.

We and others have recently been examining how effective labs are at achieving various goals. What we've found is that traditional labs fall way short of achieving a frequent goal. The surprising—even shocking—results make a strong case for reexamining what lab courses are for and how they are taught. Fortunately, certain goals and teaching methods have shown good evidence of effectiveness.

At many institutions, introductory labs are aligned with the introductory lecture course. The intended goal is to enhance student understanding of lecture content, and the underlying rationale is that students will better understand the physics if they conduct experiments and see for themselves how the physics principles work in real life. To see whether that was true, we took advantage of courses in which the lab component was optional. That way, we could compare content mastery of the students who

Natasha Holmes is an assistant professor of physics at Cornell University in Ithaca, New York. **Carl Wieman** is a professor of physics and education at Stanford University in California.





took the lab with those who did not. We considered the final exam to be a reasonable measure of content mastery.

Our investigation turned out to be more complicated than looking at exam scores, though. Students usually decided whether to take the lab based on their major or other requirement. For example, many medical schools require a physics course with a lab. Consequently, the populations of students who did and did not take the lab differed in several ways. We could detect those differences in tests of students' previous knowledge of the physics covered in the course, when such data were available.

However, we realized that we could correct for the difference between the two groups by using the fact that not all course material had an associated lab activity. We could normalize an individual student's performance on questions whose target content was reinforced by a lab activity by using the score of the same student on questions whose target content lacked an associated lab activity. That step gave us a "mean lab benefit" for the course. More precisely, the benefit was obtained by first calculating each student's fractional score (normalized to 1) on the lab-related questions minus their fractional score on the non-lab-related ones. We next calculated the average of those differences for each group of students (those taking and those not taking the lab), and then calculated the difference between the group averages.

If the lab provides the intended learning benefits, then the students who took it should have a larger difference between their lab-related and non-lab-related scores than those students who did not. How much larger? Given that labs in the courses we examined occupied 30% of the total instructional time (two out of six contact hours per week), it seems reasonable to hope for the boost to student learning on the lab-related questions to be about 30%. When we carried out the analysis of three courses at Stanford University, where we both worked at the time, we were surprised to find no difference and hence no measurable benefit.[1]

Along with collaborators Jack Olsen and Jim Thomas, we then extended the analysis to six other introductory physics courses at two other institutions that varied widely in size, geographical location, selectivity, research intensity, and student demographics. The courses were also quite different, but the labs all shared the primary goal of supporting the learning of content in the associated lecture courses. Taking the labs was optional throughout.

As shown in figure 1, the results are striking. The nine lab courses covered both mechanics and electricity and magnetism; and they used algebra- and calculus-based approaches. They were offered at 3 institutions, taught by 7 instructors, and taken by nearly 3000 students. Despite that variation, the results were the same across all nine courses. With a high degree of precision, there was no statistically measurable lab benefit.[1,2] None of the mean effects was larger than 2%; statistically, they were all indistinguishable from zero.

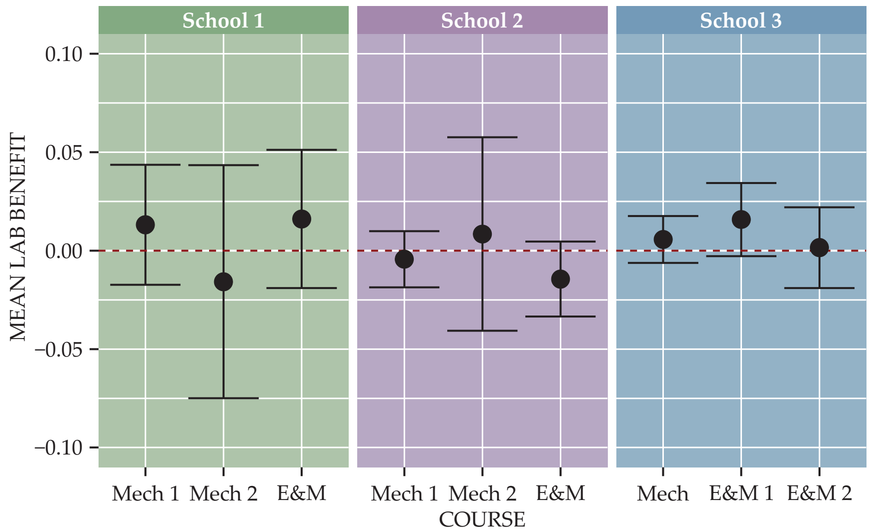


FIGURE 1. THE MEAN LAB BENEFIT at three institutions shows no measurable impact on student course performance from enrolling in the associated lab course. Mean lab benefit was obtained by first calculating each student's fractional score on the lab-related questions minus their fractional score on the non-lab-related ones. Next, the average of those differences was calculated for each group of students (those taking and those not taking the lab). Mean lab benefit is then the difference between the group averages. Error bars represent standard error of the mean. (Adapted from ref. 2.)

Leaving no stone unturned, we carried out similar analyses for midterm exams. Conceivably, any differences could be washed out by the time of the final. We also restricted our analysis to exam questions that required only conceptual reasoning and no quantitative calculations. Conceivably, labs could help students grasp the concepts. In all cases, we came up with the same null result for the lab benefit.

One can argue that labs might be achieving other educational goals that are not being measured. We cannot rule out the possibility. But given the resources devoted to such lab courses at many institutions and given the fact that the courses we studied, and similar ones at other institutions, had the express goal of reinforcing classroom instruction, that argument seems inadequate.

Labs that focus on enhancing what students learn in lectures have additional problems, as revealed by data gathered by Heather Lewandowski and collaborators at the University of Colorado Boulder on students' attitudes toward labs and experimental physics. The short, online Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS) asks students to rank how much they agree with various statements about physics experiments. "I don't need to understand how the measurement tools and sensors work in order to carry out an experiment" is one of the statements. Students' responses are scored based on how they compare with those of practicing physicists (who disagreed with the sample survey statement).

Lewandowski and Bethany Wilcox have compared thousands of students' scores on the E-CLASS across dozens of US institutions. They found that students in lab courses whose primary aim was to teach physics concepts as opposed to experimentation skills came away with beliefs that were less expert-like than the ones they held when they started.[3] Labs characterized as "guided" as opposed to open-ended also showed similar negative effects.[4]

Looking at student thinking

To try to understand why those lab courses were failing to meet their intended aims and what might be done to make them more effective, we turned to the basic method underlying most physics education research: looking at student thinking. We explored that mental terrain through an extended set of interviews with focus groups with 32 students who carried out undergraduate research over a summer semester; all the students had already taken the introductory physics lab sequence.[5] We asked them some general questions about their experiences in research, such as What are you enjoying the most? What are you learning? What's been frustrating? We then asked them to make comparisons with their lab course experiences. Lastly, we went back and characterized the discussions according to the number of times students mentioned carrying out or not carrying out various thinking, or cognitive tasks, involved in doing authentic experimental physics. Such tasks include defining goals, testing equipment, and presenting results[6] (see figure 2).

The only thinking the students said they did in structured and content-focused labs (the kind in our study of nine courses) was in analyzing data and checking whether it was feasible to finish the lab in time.[5] Although the finding may seem surprising at first, if you break down the elements of a typical lab activity, you realize that all the decision making involved in doing experimental physics is done for the students in advance. The relevant equations and principles are laid out in the preamble; students are told what value they should get for a particular measurement or given the equation to predict that value; they are told what data to collect and how to collect them; and often they are even told which buttons to press on the equipment to produce the desired output.

The lack of decision making also explains the results in figure 1. Although the students are going through the motions of physics experimentation, their brains are not engaged in the process, and there is little need or reason to think about the physics content involved. That mental effort is made by instructors beforehand when they design the experiment and when they think about the research questions and how to test them. Our research suggests that instructors are erroneously assuming the students will go through a comparable thought process as they follow the instructions in the lab manual to complete the experiment in the allotted time.

The focus group interviews also provide some guidance on the challenges of achieving the labs' ostensible goal of teaching physics and some ideas for other educational goals that labs might better achieve. Students explicitly talked about doing almost all of the specific cognitive tasks needed for typical experimental physics research as part of their summer research projects (referred to here as undergraduate research experiences; UREs). Conversely, they frequently mentioned why they could not do the same in their structured lab courses.

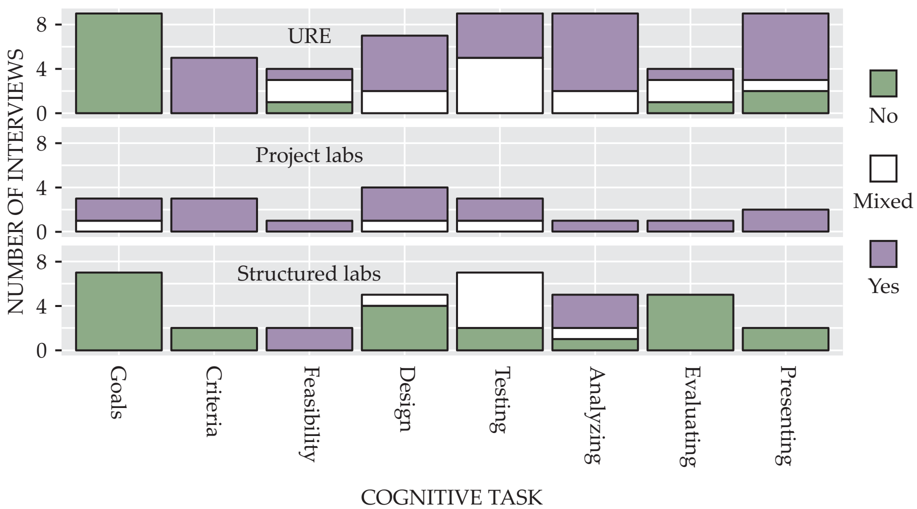


FIGURE 2. THE NUMBER OF FOCUS GROUP INTERVIEWS in which students explicitly described engaging in ("Yes") or not engaging in ("No") various cognitive tasks associated with experimental physics in their undergraduate research experiences (UREs), project labs, and structured lab courses. The tasks are setting goals, establishing success criteria, determining feasibility, designing the experiment, testing equipment, analyzing data, evaluating results, and presenting results. "Mixed" refers to the case in which some students said they did engage in that activity and others said they did not. (Adapted from ref. 5.)

For their summer projects, the students did not develop their research goals, which were laid out for them in advance by their advisers. Nevertheless, the students emphasized, usually in positive terms, the freedom they had to figure out how to achieve their part of the project. When they inevitably got stuck, they could try to fix the problem, take a break and come back with a new idea, ask for help and feedback, and so on. As an embedded participant, they attended group meetings and got to observe the scientific process unfold, while also being in control of a small piece of it.

In interviews, students repeatedly mentioned the importance of having the opportunity to make decisions on their own. Also important was having the time both to reflect on those decisions and their outcomes and to fix and improve the experiments iteratively. They recognized how those two features—making decisions and taking time—contributed much of the value they saw in their research experiences. They could also see that their structured lab courses lacked those features.

Most of the students we interviewed had also taken at least one project-based lab course that was entirely open ended and focused on engaging them in authentic research experiences. In those courses, students designed, built, and carried out their own experiments to answer their own research questions. Like UREs, such courses were not intended to support learning the lecture content. But unlike UREs, the courses shared many of the same constraints as the structured, concepts-focused labs. Time in the lab and with instructors was limited, and the content of the experiments was related to the associated courses. In our interviews, students brought up project labs much less often than structured labs or UREs, but the cognitive tasks the students described were close to those in the UREs, as was the opportunity for agency and iteration. That finding is further supported by the E-CLASS studies indicating the value of having an element of open-endedness in labs. Open-endedness implies opportunities for student decisions and not rigid cutoffs for completion, though the spectrum of open-endedness makes it unclear what level is optimal.

The issues raised above reveal a fundamental constraint on the ability of labs to support the learning of content in an associated lecture course. It is likely that students did learn the physics content associated with their experiments, either in the project lab or URE, but that physics content formed only a narrow slice of a typical lecture course. To cover a reasonable fraction of concepts in an associated course, a lab must be highly structured like a cookbook. Only then can students complete it quickly, obtain the desired result, and go on to do many other experiments related to the rest of the material in the lecture course. But labs that can be completed quickly do not require students to do much of the thinking that they need to learn the content.

As first pointed out to us by Douglas Bonn at the University of British Columbia, the ultimate problem with trying to use labs to teach physics content is not that it is impossible. Rather, the labs take far more time—more overhead if you like—than do other methods for teaching the same content.

Studio physics is one fairly radical alternative. Instruction takes place in a specially designed space around a series of small experiments. Students move continuously between the different types of activities—doing experiments, consulting online resources, talking to instructors—so the activities are mutually supportive. Although studio physics has been shown to improve students' understanding of content, the impact of the hands-on activities has not yet been disaggregated.

Thinking like a physicist

Another alternative, which we developed with Bonn, is more suitable for large, stand-alone introductory lab courses. The

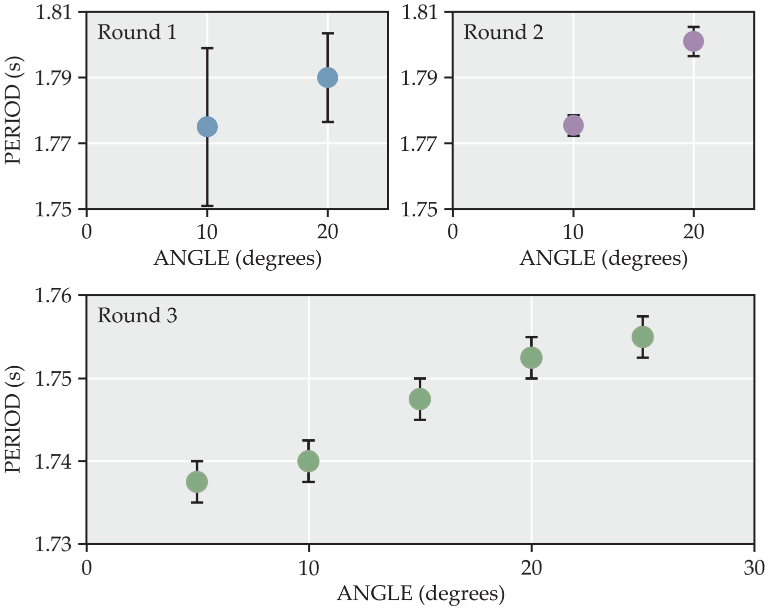


FIGURE 3. ONE STUDENT'S MEASUREMENTS of the period of a pendulum as a function of amplitude. The first round of measurements shows the expected null effect. A second round that includes more trials and measures a greater number of swings illuminates the distinguishability of the two measurements. The third round includes measurements at more amplitudes and suggests the second-order quadratic deviation from the small-angle approximation.

approach involves abandoning the goal of teaching content and instead aiming to teach important aspects of physics thinking for which labs are uniquely effective. Those aspects relate to the process of scientific experimentation, such as formulating hypotheses and testing them by collecting data, figuring out how to improve the quality of data, using data to evaluate the validity of models, and deciding on suitable criteria for such evaluation.

To learn those process skills, students must practice and learn the thinking that physicists do. For students to succeed, guidance must be sufficient but not too constraining. Students must still have a sense of agency to make their own decisions and enough time to go back and change something if it doesn't work. Overcoming obstacles and learning from failure are vital skills for every experimental scientist. To acquire them, students need significant support as well. A small number of examples of instructional lab pedagogies have incorporated a balance of agency and support the aim of teaching the skills of an experimentalist and have demonstrated evidence of their effectiveness. We discuss two examples below.

Our own structured quantitative inquiry labs (SQILabs) explicitly focus on iterative experimentation and decision making. The goal is to develop students' quantitative critical thinking skills,[7] defined as the ability to use data to test and evaluate models, explanations, and methods. Students must therefore understand experimental uncertainty and its implications in such tests. In a SQILabs activity, students are given a relatively constrained experimental goal and setup, but they decide how they conduct the experiment and interpret the data.

In one SQILabs activity, students compare the period of a pendulum when released from two angles of amplitude.[8] They must make many experimental decisions, such as how many repeated trials to take and how many times to let the pendulum swing for each measurement. Most often, students' initial investigations appear to confirm the prediction from the lecture class or textbook: The period does not depend on the amplitude. Students are then instructed to iterate—that is, to find and implement ways to improve their measurement, with emphasis on reducing measurement uncertainties. Once again, they must decide on their own how to do so.

As the students' measurement uncertainty decreases, their results begin to expose a subtle difference between the periods with the two different amplitudes. Once they have encountered the surprising result, students are instructed to hypothesize possible explanations and to design and carry out follow-up experiments. Is the model wrong? Is air resistance responsible? Does the effect appear at smaller or larger amplitudes? Is it a measurement mistake? When the students measure the period across several amplitudes with sufficient precision, they discover the second-order quadratic behavior that deviates from the small-angle approximation, but only if they properly understand and trust the quality of their data. (See figures 3 and 4.)

The pendulum experiment and others like it engage students in high-level critical thinking. They connect the quality of the data collected to the depth of the physics that can be explored.[7] Students in SQILabs also retained that physics thinking in their lab course the following year, despite its traditional, highly structured design. By contrast, the more traditional version of the lab experiment would not yield data of sufficient precision for the students to discern the second-order behavior; any discrepancies would be routinely dismissed as "human error," and if they wondered about a possible discrepancy, they would have no time to go back and examine it. In an interview, a SQILabs student reflected on how this approach helped in grasping an essential aspect of physics:

> In physics, there's lots of simplifications and approximations and things that we can ignore. When we do the experiments ourselves, we can see why physicists would do that.

The process of doing a SQILabs experiment even led students to consider the ethical and philosophical issues in scientific research. Here's a comment from another interviewee:

> The pendulum experiment we did at the beginning of the year, I think that really made a mark on me. Because I went in there expecting it [the period at 10 and 20 degrees] to be the same, because that's what I was taught. And then, when you finally figure out that, "Oh, it's supposed to be different," and then I was like, "Oh! I probably shouldn't be doing experiments with bias going in."

Furthermore, attitudes and beliefs toward experimental physics did not show the typical declines observed in introductory lab courses.[9]

A second example of a process-focused introductory lab is

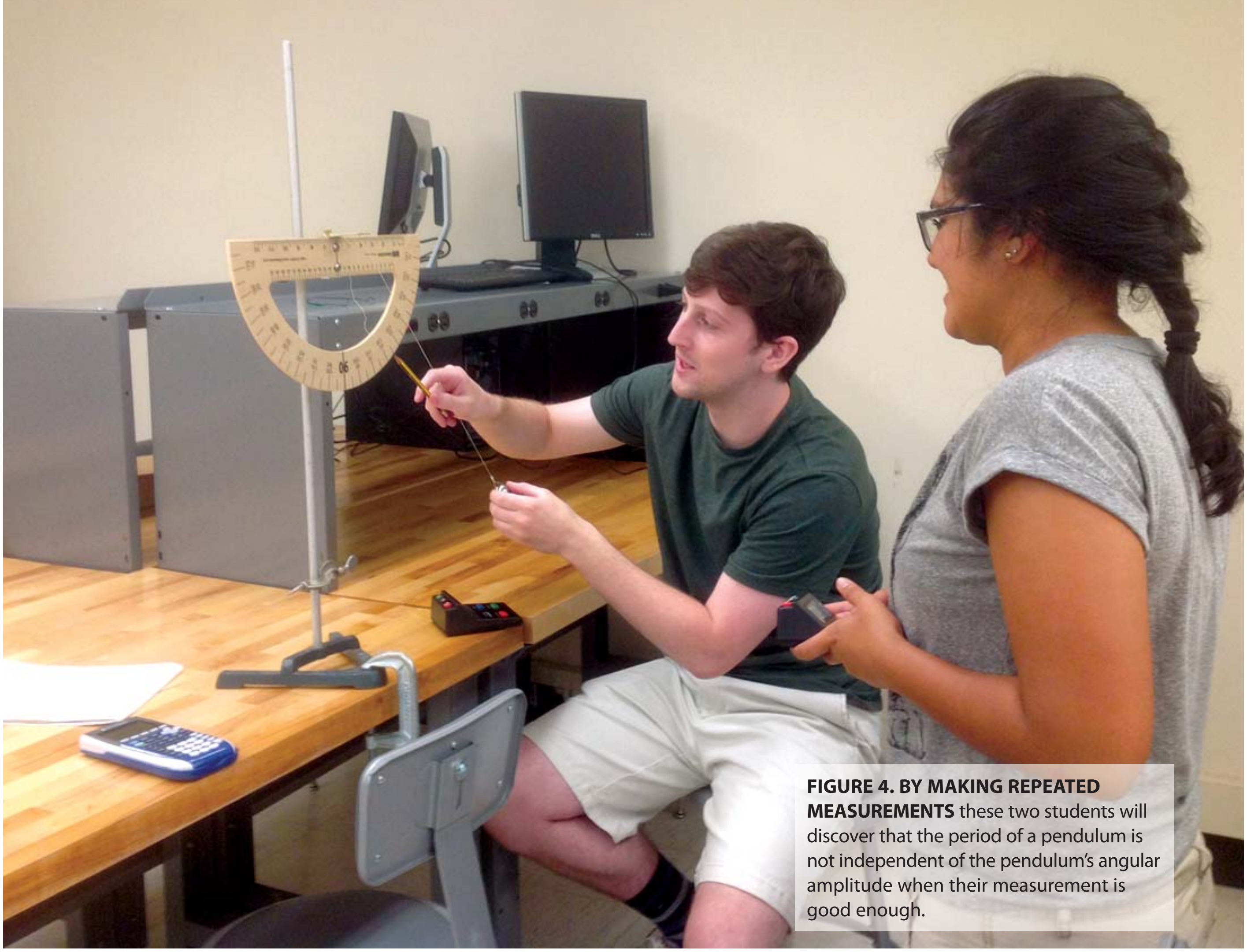


FIGURE 4. BY MAKING REPEATED MEASUREMENTS these two students will discover that the period of a pendulum is not independent of the pendulum's angular amplitude when their measurement is good enough.

the Investigative Science Learning Environment (ISLE) designed by Eugenia Etkina and coworkers at Rutgers University.[10] In an ISLE course, students work in groups to develop explanations of new and surprising phenomena and then design experiments to test those explanations. The phenomenon makes its first appearance in the form of a demonstration selected by an instructor or a simple experiment carried out by the students in groups. In one example, the initial phenomenon is a light cone produced when a green laser shines through a tank of water onto a piece of paper.[11] Students work together to develop a list of possible hypotheses that explain the phenomenon (for example, that the paper reflects the light back into the tank or that the interface between the glass in the water causes the shape). They then design experiments that can test whether each hypothesis can uniquely account for the phenomenon. As the groups carry out each experiment, they gradually winnow the list of plausible hypotheses. The process is iterative in that students' experiments evolve to construct the physical model. In subsequent activities, students design and perform new experiments that not only apply and build on their newly developed explanations but also lead them to observe new phenomena, starting the cycle again.

As students participate in ISLE labs, they acquire several scientific abilities,[12] such as finding qualitative and quantitative patterns in phenomena, proposing multiple models or explanations, designing and carrying out experiments to test them, evaluating assumptions and uncertainties, representing data and evidence, and making comparisons between uncertain measurements. Instructor-guided discussions engage students creatively in the scientific process: Any new idea must be testable, and the tests must isolate the desired effect. Most of the skills overlap those expressly developed in SQILabs, and there is good evidence that both approaches are effective. Although ISLE and SQILabs differ in their settings, some of the goals, and in the details of implementation, they both focus on the experimentation process: They both allow students to make decisions, act on those decisions, and iterate to improve their data and models.

In both SQILabs and ISLE, the instructional goals are carefully chosen to be limited and realistic, and the lab activities are carefully structured so that students can practice the thinking, reasoning, and decision-making processes involved in experimentation. Although both approaches are structured, they differ from a traditional "cookbook lab" in what is structured. In a traditionally structured lab, students are told what to do and how to do it. Decisions are made for them. In the ISLE and SQILabs examples above, students are told what decisions need to be made, but it is up to them to make the decisions and act on them. The structure of ISLE and SQILabs activities also explicitly prompts students to reflect on and evaluate the outcome of their decisions and to act on that evaluation. The prompts should decrease with time, so that by the end of the course the students can make decisions without them. They get to create and explore novel explanations and experimental designs—a degree of creativity that is missing from conventional labs. In interviews, students made it clear that the sense of both agency and creativity contributed greatly to their enjoyment and motivation.

Student interviewees also revealed that when they had to obtain a prescribed outcome in a traditional lab—say, measuring the value of the acceleration due to gravity—they became frustrated when things did not work out as planned. Often, they resorted to massaging data to get the expected result.[5] Providing experimental goals for which the outcome is either unknown or unexpected (for example, the second-order effect of a pendulum's amplitude dependence) moves the focus away from what students measure and toward how they measure it. The new focus reduces student frustration and instills more scientifically authentic and desirable behavior. It also provides opportunities to talk about the nature of measurement (What is a "good" measurement and how do we know?) and about argumentation from evidence (What can we conclude about the measurements given their quality?).

The shift in goals and structure that we advocate necessitates reallocating the way time is spent in labs. Because physics experiments rarely work out as expected, and are virtually never done as well as they could be the first time, students need time to troubleshoot, revise and test models, and try new approaches. Supporting agency and iteration also means students need to spend more time making their own decisions and learning from their choices.

Spreading individual lab activities over multiple sessions is an easy way to provide the necessary time. For example, the pendulum lab in SQILabs described above takes place over two lab sessions. First, students learn about uncertainty, standard deviation, and comparing uncertain measurements. Next, they iteratively measure the period from the two amplitudes. Finally, they design and conduct experiments to explain the measured difference in periods. Of course, having multiple-session experiments means that students will complete many fewer experiments than is the case in highly structured lab courses, but they achieve the desired educational goal—learning to think and behave like a physicist—far more effectively.

Investing in effectiveness

Any course transformation will require an initial investment, but the cost of implementing the design principles can be modest. The labs can be done with the same space and, depending on the choices made, with much the same equipment. One unavoidable cost is for training teaching assistants, as they need to fill quite a different role than they do in traditional introductory labs. They also will likely need to spend more time evaluating student lab work, as they have to shift from simply assessing a product (Did the students get the right answer? Did they use the correct format in graphing results?) to instead providing feedback on the students' processes. How students conduct the experiments may vary widely. Consequently, teaching assistants' feedback must be responsive to a range of students' ideas. To reduce the total grading time, both ISLE and SQILabs moved to evaluating the lab notes of each group rather than of each individual's and to using carefully constructed rubrics to streamline grading.

At the University of Illinois at Urbana-Champaign, instructors are implementing ISLE-inspired labs with modifications to reduce the demands on classroom space, equipment, and instruction time and to allow many more students to participate without increasing costs. Mats Selen and Timothy Stelzer have developed the IOLab, a pocket lab apparatus that has sensors for measuring just about any introductory physics lab experiment. Students can take it home to explore an aspect of the target phenomenon or measurement process that will support the in-class experiments. They then come to class ready to discuss their results, develop new skills, and build on their initial experiments in an ISLE-like format. Preliminary studies are showing that the IOLab course structure improves students' attitudes and beliefs about experimental physics compared with those of students in control groups in traditional concepts-focused labs.[13]

Measuring the effects of new approaches to teaching introductory labs is crucial and nontrivial. Although research-based diagnostic assessments for measuring students' conceptual physics understanding are abundant, similar tools for measuring lab skills are scarce. The E-CLASS measures student attitudes and beliefs about physics experimentation skills and concepts. The Concise Data Processing Assessment[14] and the Physics Measurement Questionnaire[15] are designed to evaluate students' understanding of uncertainty and data analysis. Research-validated rubrics can also be used to assess students' scientific abilities.[12] We are currently validating the Physics Lab Inventory of Critical Thinking, which assesses students' proficiency with evaluating experimental designs, data, and models.[16]

Introductory labs offer unique opportunities to teach experimentation, reasoning, and critical thinking skills, which can be of value to all students, regardless of their major. We argue that to learn such skills, the lab experience must provide students with opportunities for decision making and with the time to learn from their decisions. Rather than being seen by students as pointless and frustrating hoops that have to be jumped through, introductory physics labs can instead offer rewarding intellectual experiences.

We thank Peter Lepage, Heather Lewandowski, Katie Ansell, and Eugenia Etkina for feedback on the manuscript.

REFERENCES

1. C. Wieman, N. G. Holmes, *Am. J. Phys.* **83**, 972 (2015).
2. N. G. Holmes et al., *Phys. Rev. Phys. Educ. Res.* **13**, 010129 (2017).
3. B. R. Wilcox, H. J. Lewandowski, *Phys. Rev. Phys. Educ. Res.* **13**, 010108 (2017).
4. B. R. Wilcox, H. J. Lewandowski, *Phys. Rev. Phys. Educ. Res.* **12**, 020132 (2016).
5. N. G. Holmes, C. E. Wieman, *Phys. Rev. Phys. Educ. Res.* **12**, 020103 (2016).
6. C. Wieman, *Phys. Teach.* **53**, 349 (2015).
7. N. G. Holmes, C. E. Wieman, D. A. Bonn, *Proc. Natl. Acad. Sci. USA* **112**, 11199 (2015).
8. N. G. Holmes, D. A. Bonn, *Phys. Teach.* **53**, 352 (2015).
9. L. E. Strubbe et al., in *2016 Physics Education Research Conference Proceedings*, D. L. Jones, L. Ding, A. L. Traxler, eds., American Association of Physics Teachers (2016), p. 340.
10. E. Etkina, *Am. J. Phys.* **83**, 669 (2015).
11. E. Etkina, G. Planinšič, M. Vollmer, *Am. J. Phys.* **81**, 815 (2013).
12. E. Etkina et al., *Phys. Rev. ST Phys. Educ. Res.* **2**, 020103 (2006).
13. K. Ansell, M. Selen, in ref. 9, p. 36.
14. J. Day, D. Bonn, *Phys. Rev. ST Phys. Educ. Res.* **7**, 010114 (2011).
15. S. Allie et al., *Int. J. Sci. Educ.* **20**, 447 (1998).
16. N. G. Holmes, C. Wieman, in ref. 9, p. 156.