

Measuring the impact of introductory physics labs on learning and critical thinking

Carl Wieman^{1,2} and N.G. Holmes¹

¹*Stanford University, Department of Physics,
382 Via Pueblo Mall, Stanford, CA, 94305*

²*Stanford University, Graduate School of Education,
485 Lasuen Mall, Stanford, CA, 94305*

Abstract. Our recent study showed that two lab courses, whose goals were exclusively to reinforce material developed in the lecture courses, do not have any impact on exam performance at the 2% level. In this study, we replicated this analysis with a modified version of one of these lab courses whose goals also included modeling, designing experiments, and analyzing and visualizing data. This modified course used the same sets of apparatus as the previous version, but changed the pre-lab and in-lab activities to focus on developing and testing models with data. The study evaluated the impact of these additional goals and activities, comparing performance with students in the same course who did not take the lab. We found that students taking the lab still performed no better or worse on the final exam than students who did not take the lab. We also observe the critical thinking behaviors that were the new goals of the lab.

PACS: 01.40.G-, 01.40.-d, 01.50.Qb, 01.55.+b

I. INTRODUCTION

The American Association of Physics Teachers [1] recently released recommendations for learning goals for the undergraduate physics lab curriculum. The document focuses on six areas: modeling, designing experiments, developing technical and practical skills, constructing knowledge, analyzing and visualizing data, and communicating physics. In practice, actual lab courses focus on a wide variety of goals, including developing conceptual understanding of physics content [2]. Seldom do they consider which goals can be uniquely and optimally achieved in labs. Also, there has been little work examining what students are actually learning in introductory labs [2]. In a recent study, we examined the effectiveness of intro labs at meeting the traditional goal of reinforcing physics concepts and content presented in lecture [3]. We found that a lab course with explicit learning goals focused on reinforcing the lecture content, and coordinated with the lecture course, did not have any impact on exam performance with an uncertainty at the 2% level. This was true for two different calculus-based introductory physics courses, a mechanics course and an electricity and magnetism course. This evaluation supports the ongoing discussion in the lab community surrounding the goals and effectiveness of the introductory physics labs.

We recognize that there are many physics courses where the hands-on laboratory portion of the course is embedded within the lecture portion (such as the studio or workshop physics [4-5], SCALE-UP [6], or ISLE [7] approaches). In these approaches, there are clear improvements in students' learning of the content material over traditional lecture format, but given the integrated format of the courses, it is unclear how or whether the hands-on activities contribute to that learning. In addition, many large universities do not

have the resources to support these integrated courses. It is important, therefore, to evaluate how or whether stand-alone lab courses contribute to student learning.

Given the results of the initial study [3], we redesigned the lab course associated with the electricity and magnetism course at the same institution in the subsequent year with a new set of learning goals and restructured pre-lab and in-lab activities accordingly. The redesigned lab course used the same physical apparatus as the previous version of the course, so both courses involved investigating the same underlying physics concepts. The redesigned lab, however, included explicit goals about evaluating data and models. Our research question for this paper was to evaluate whether shifting the focus of the lab goals and activities affected students' performance in the lecture course, as measured by the final exam. We also provide preliminary evidence that the modified lab course engaged students in high-level reasoning and modeling behaviors.

II. MODIFIED LAB COURSE

The goals and activities of the modified lab course focused on comparing and interpreting measurements, models, and data and evaluating and refining models based on data and assumptions about the system. The redesign was inspired by an introductory lab framework that has students reflect on their data and results and use that reflection to iterate to improve their results or knowledge of a system [8-10]. Through this iterative experimentation process, students were seen to engage more deeply with the physical and measurement models involved in experimentation [11], even leading them to reexamine the assumptions about the system and refine physical models seen in lecture, based on their measurements.

To redesign the course, the existing lab apparatus were evaluated for measureable model limitations, including

what level of precision and analysis was necessary for students to observe such limitations. The apparatus and original experiments were quite conventional, covering Coulomb's law, Faraday's law, simple circuits, etc. and are fully listed in [3]. Only minor adjustments to the order of the experiments were needed to incorporate the goals about evaluating data and models in a productive sequence. For the first half of the course, pre-lab activities focused on introducing data analysis concepts and tools. These had previously not been part of the activities associated with this sequence of labs. The in-lab activities focused on giving students deliberate practice with those tools and concepts, exploring them in a physical context.

For the second half of the course, the activities focused more explicitly on modeling and testing models. Pre-lab activities used sequences of questions to direct students through the underlying physical models. For example, a resistor-capacitor experiment used the pre-lab to explore why, physically, a capacitor would discharge according to exponential decay, rather than another model. The in-lab activity, correspondingly, involved using measurements to verify the exponential model. Another experiment explored the magnetic field from a loop of current. The pre-lab focused on the idealized model, while the in-lab activity focused on leading students to use their data to identify invalid assumptions and limitations of the model. In particular, students discovered issues with measurements taken off-axis due to the fringing effects from the loop.

III. METHODS

Participants were 443 students in an introductory electricity and magnetism course at a large, elite university who completed the final exam for the lecture course. A subset of those students ($n=126$) was also enrolled in the associated lab course. Students in the two groups were not equivalent in background since enrollment in the lab course was encouraged for different majors. The lecture course involved 3 hours per week of in-class lecture plus 2 hours per week in discussion sections led by graduate teaching assistants (TAs). The lab course involved weekly 2-hour labs with approximately 15 students led by a graduate TA. A physics education researcher designed the pre-lab and in-lab activities and coordinated weekly TA training meetings.

At the beginning of the lecture course, students in both groups (the lab group and the non-lab group) completed the Conceptual Survey on Electricity and Magnetism (CSEM) [12] during the first discussion sections. This survey was used to compare the two groups on entering the course.

Student scores on 20 multiple-choice items on the final exam in the lecture course were used as a performance measure of content learning. Not all items were related to the physics principles discussed in the lab. This structure allows us to compare students who did and did not take the lab on content relevant and not relevant to the lab. Two independent raters coded whether each multiple-choice

question was related to or not related to a lab experiment. The descriptions of specific activities in the pre-lab and in-lab activities were used for this evaluation. One of the raters did not look in detail at the lab activities and the two raters agreed on 13/20 items. The remaining items were discussed in the context of a more detailed examination of the student activities in the labs, and the researchers quickly reached consensus on all 20. From this coding, 7 out of 20 questions were coded as being related to the lab content, with 13 questions not related to the lab content.

A ratio was calculated for each student comparing their average score on the lab-related questions with their average score on the non-lab related questions, as in the original study [3]. If the ratio is greater than one, then the student performed better on the lab-related questions than the non-lab-related questions. If the ratio is less than one, the student performed better on the non-lab-related questions than the lab-related ones. Finally, if the ratio is equal to one, then they performed equally well on both types of questions. We then average those ratios across the students who did (or did not) take the lab course.

An average ratio that is greater for the lab students than the non-lab students indicates that taking the lab improved the students' learning of those topics covered by the lab. An average ratio that is greater for the non-lab students suggests that the labs did damage to students' understanding of the lab content. Average ratios that are the same for both groups implies that the labs had no added benefit to students' understanding of the content, as was found in the previous study [3]. We also evaluate the lab benefit difference: the difference between the two groups on the difference between their scores on the lab and non-lab related questions.

Based on our previous work, the most plausible hypothesis is that there would be no difference between the two groups on either of these measures, but there are two plausible alternatives. First, the shift in lab goals and activities distracted students and shifted their focus, negatively interfering with learning the lecture material. Second, the shift in lab goals and activities enhanced the students' understanding of the associated lecture materials. In [3] we presented statistical arguments showing the lack of correlation between responses to the different exam questions, and hence why it is implausible that taking the lab would improve performance on both lab and non-lab related questions.

We also assessed students' performance evaluating data and models. During the 6th week of the course, students conducted an experiment exploring the magnetic field created by a current-carrying loop. They measured the deflection of a compass at the center of the loop as a function of the current in the loop and used this data to measure the magnitude of Earth's magnetic field. The model assumes that all measurements are exactly at the center of the loop, which is increasingly flawed as the angle increases. This resulted in non-linear behavior in their

graphs at large currents and angles. In the modified lab course, this investigation had three parts.

Part 1 involved taking and interpreting a high-quality data set across the range of angle deflections. They were encouraged to use drawings of the magnetic field from the apparatus to explain unexpected results. In Part 2, they were explicitly asked to evaluate the assumption that the compass was at the center of the coil. In part 3, they were told to design and carry out a measurement that would reduce the effects of these invalid assumptions to obtain a high quality measurement for the magnitude of the local magnetic field. (Not the Earth’s true field, since the experiment was conducted in the basement of a physics building and they were only measuring the horizontal component of the field.)

To evaluate students’ performance on this activity, we coded the quality of students’ reflection on data and models using a scheme described in [9,10]. For both Part 1 and 3, we evaluated how students interpreted their data on a scale from 1 (only presented results with no interpretation) to 4 (synthesized multiple components of the results and evaluated them). For all three parts, we also coded whether students evaluated the given models and assumptions, either identifying a disagreement or by interpreting it physically.

IV. RESULTS

On average, students in the course correctly answered 14.91 ± 2.51 ($M \pm SD$) out of the 20 final exam multiple-choice questions. Scores were normally distributed (skewness = -0.46). We did basic statistical tests of the reliability and validity of these questions as an assessment instrument. We calculated the Kuder-Richardson Formula 20 (KR-20) for the full set of 20 multiple choice items, which is analogous to Cronbach’s alpha, but for dichotomous data. This was found to be $KR-20 = 0.63$ for the exam, which is an acceptable range for an internally consistent (reliable) test, measuring multiple constructs. Correlations between individual questions and the total score ranged from 0.15 to 0.51, with only two questions having $r < 0.2$. This suggests the test items reasonably discriminate between high and low performing students.

All assumptions for performing an independent-samples t -test on the ratios were met. Three such assumptions involve the study design and measurements: the dependent variable is continuous; the independent variable consists of two independent groups (a lab group and a non-lab group of students); and there is independence of observations (lab students’ performance on the final exam does not effect the non-lab students’ performance). The final three assumptions relate to the characteristics of the data: there were no significant outliers as evaluated on a histogram of students’ ratios for each group; the student ratios were normally distributed for both groups; and there was homogeneity of variances (see data in the results section).

As in Ref. [3], there was a significant difference between the two groups on the 20 multiple choice items on

the final exam (Table 1), with the lab students outperforming the non-lab students: $t(441) = 3.64$, $p < .001$, 95% confidence interval for the difference is [0.44, 1.46]. As noted above, this was not surprising as the students in each group tended to come from different majors. These differences were also reflected in the difference between the two groups on the CSEM pre-test, with the lab students ($M=15.96$, $SD=5.62$) outperforming the non-lab students ($M =14.42$, $SD =5.13$): $t(395) = 2.69$, $p = .008$, 95% confidence interval of the difference is [0.41, 2.66].

There was no significant difference between the two groups on their ratio scores (average lab-item score over non-lab-item score, see Table 1): 95% confidence interval of the difference is [-0.03, 0.07]. We also compared the two groups on the average difference between lab related and non-lab related questions. This lab benefit difference between the two groups was of similar magnitude (at the 2% level), and negligible, as that in the previous study. These results demonstrate that the students who took the lab came in and left the course with higher ability in the course content. It is clear, however, that taking the lab course does not interact with this gap and raises questions about what the lab is achieving.

TABLE 1. Student scores on elements of the final exam by group as $M \pm SD$

| | # items | Non-lab students (n=317) | Lab students (n=126) |
|---------------------------------------|---------|-------------------------------------|-------------------------|
| Overall score | 20 | 73%±12% | 78%±13% |
| Lab related questions | 7 | 72%±18% | 78%±19% |
| Non-lab related questions | 13 | 74%±13% | 78%±13% |
| Average Ratios | | 0.99±0.26 | 1.01±0.25 |
| Difference between ratios | | 0.02±0.03; $t(441)=0.72$, $p=.472$ | |
| Lab benefit difference between groups | | 0.016±0.018 | |

Next we assess students’ behaviors evaluating data and models on the Earth’s magnetic field experiment described in Section III. First, we measure their reasoning about data. From Fig. 1, we see most students in Part 1 of the activity reflecting on their data at high levels (level 4), but lower levels in Part 3. This difference may be because the disagreement in Part 3 was evidenced only by a high quality data set, so fewer students noticed it. Students tend to reason at lower levels when they obtain agreement [9,10]. This is also seen in how they evaluated models (Fig. 2).

We see fewer students evaluating the model in Part 3 of Fig. 2 than in Parts 1 or 2. This is consistent with students’ lower level reasoning in this part and the fact that the disagreement was less evident in Part 3. Even though all three parts had explicit prompts to evaluate, the focus of the prompts in Parts 1 and 3 was to evaluate data, while in Part 2 it was to evaluate the model. When students were explicitly told to evaluate the model (Part 2), they did.

To provide context to these results, we compare to the same analysis in related work [9,10]. Less than a quarter of students evaluated models in this way at the start of a similar-level course and at the end of a traditional course. At the end of a course with similar goals and structure as the current modified lab course, over 75% of students evaluated models this way without instruction to do so, compared to 60-70% of students in Parts 1 and 3 here.

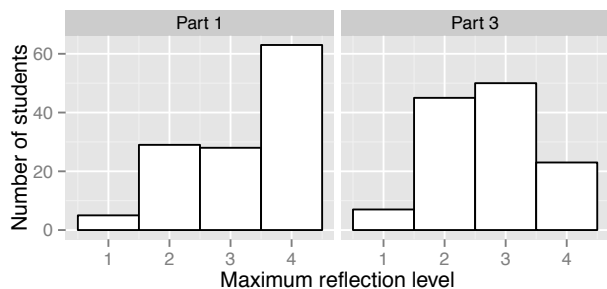


FIG 1. Students' levels of reasoning about their data.

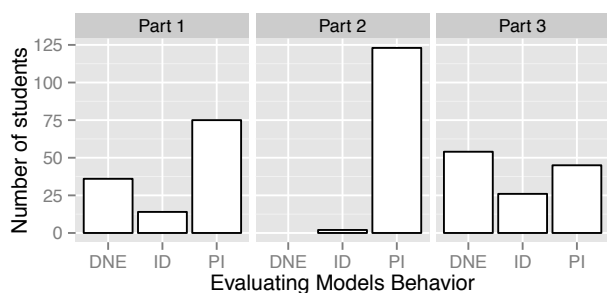


FIG 2. Respective numbers of students: not evaluating the models (DNE); identifying disagreements between their data and a model (ID); and interpreting it physically (PI).

V. DISCUSSION

In this study, we replicated the methodologies of a recent study that evaluated student learning of lecture course content in a lab course that exclusively aimed to reinforce the lecture course content, but provided no added value [3]. In the subsequent year, the lab course was redesigned to incorporate learning goals and activities focused on evaluating data and models. This study found

[1] American Association of Physics Teachers Technical Report, 2014.
 [2] S.R. Singer, M.L. Hilton, and H.A. Schweingruber, *America's Lab Report: Investigations in High School Science* (National Academies Press, 2005).
 [3] C. Wieman and N.G. Holmes, *Am. J. Phys.* (to be published).
 [4] P.W. Laws, *Phys. Today*, **44**, 24 (1991).
 [5] P.W. Laws, *Change*, **23**, 20 (1991).
 [6] R.J. Beichner *et al.*, *Research-Based Reform of University Physics* **1**, 2 (2007).

that this lab, again, had no impact on final exam score in the associated lecture course.

The redesigned lab course, however, offered additional learning experiences, adding learning goals about evaluating data and models. There is preliminary evidence that students were engaging in high-level and desirable critical thinking about data and models. We argue, therefore, that the redesigned lab course *did* offer added value that was complementary to the lecture course content.

While we do compare results here to previous work, one limitation of this study is the lack of a control group for the in-lab behaviors. From interviews and discussions with students, TAs, and instructors, however, we have seen no indication that students were engaging in these behaviors in prerequisite lab courses, or in previous years of the course.

Another limitation of this study is in our measurement of learning of the lecture course content, as the final exam was not evaluated for validity or reliability a priori. The measures included in the results suggest that the exam had sufficient discrimination between high and low performing students and was measuring more than a single construct. Unfortunately, the CSEM was not administered at the end of the course, so we cannot use that as an independent measure. The full set of multiple choice items on the final exam were, however, able to discern the same difference between the two groups as the CSEM pre-test. This suggests that there is some reliability that the final exam is measuring electricity and magnetism concepts.

Based on the evidence thus far, this paper, in combination with the previous paper [3], further demonstrates that a lab course that is distinct from the lecture course does not provide added value to learning the traditional course material as measured by the final exam. This was true whether the goals were entirely focused on content, or included goals about evaluating data and models. We also demonstrate that the focus on evaluating data and models produces a number of complementary learning experiences that engage students in high-level reasoning and experimentation.

ACKNOWLEDGEMENTS

We are pleased to acknowledge the assistance of Chaya Nanavati and Mark Kasevich in carrying out this work.
 [7] E. Etkina and A. Van Heuvelen, *Research-Based Reform of University Physics* **1**, 1 (2007).
 [8] N.G. Holmes and D.A. Bonn, *Phys. Teach.* **53**(6), 352 (2015).
 [9] N.G. Holmes, C. Wieman, and D.A. Bonn, *P. Natl. A. Sci.* **112**(36), 11199 (2015).
 [10] N.G. Holmes, Ph.D. thesis, University of British Columbia, 2015.
 [11] B.M. Zwickl, N. Finkelstein, and H.J. Lewandowski, *Am. J. Phys.* **82**, 876 (2014).
 [12] D.P. Maloney, T.L. O'Kuma, C.J. Hieggelke, and A. Van Heuvelen, *Am. J. Phys.* **69**, S12 (2001).