Impact of ISLE-based labs in courses with traditional lecture

Joshua Rutberg

Graduate School of Education, Rutgers University, 10 Seminary Place, New Brunswick, NJ, 08901

Marina Malysheva, Eugenia Etkina

Graduate School of Education, Rutgers University, 10 Seminary Place, New Brunswick, NJ, 08901

ISLE -based labs have been demonstrated to be an effective alternative to traditional lab activities, helping students develop scientific abilities when used in an ideal environment where lectures and recitations also reflect ISLE philosophy. In this study we used ISLE-based labs in a second-year physics course for engineering students in which the lectures and recitations remained traditional. We investigated the effect that ISLE-based labs had on student development under these conditions. We found that while students did not show as much growth as they had in previous studies under more ideal conditions, the students did show improvements in their scientific abilities over the course of one semester of instruction. We also investigated how the accuracy of feedback given to the students by the TAs related to student learning.

2019 PERC Proceedings edited by Cao, Wolf, and Bennett; Peer-reviewed, doi.org/10.1119/perc.2019.pr.Rutberg
Published by the American Association of Physics Teachers under a Creative Commons Attribution 4.0 license.
Further distribution must maintain attribution to the article's authors, cover page, and DOI.

I. INTRODUCTION

Experimental work has long been held as an essential aspect of physics education, particularly at the introductory level. While the goal of the inclusion of hands-on work may vary between institutions and instructors, one of the primary goals is often for students to develop the science process skills used by practicing scientists such as designing experiments, reasoning from data, and solving complex open-ended problems. Traditional laboratory activities, however, do little to achieve this goal [1, 2]. It has been shown that more modern approaches to instruction do meet this goal if they are intentionally designed with that goal in mind [2].

Investigative Science Learning Environment (ISLE) is a framework for creating holistic learning environments in which students are encouraged to develop their physics knowledge by engaging in processes similar to those physicists use to construct knowledge [3]. These environments are designed to reflect the practices of real scientists and engineers by engaging students in observing phenomena, coming up with potential explanations, and testing them while they are developing scientific knowledge. ISLE-based materials explicitly include observational experiments to challenge the belief that experiments serve solely as a means of testing hypotheses [4].

When used to guide introductory physics course reform, the ISLE framework has been found to improve students' scientific abilities [5]. However, these findings occurred under ideal conditions. The labs were integrated into a course with lectures and recitations and were used as a driving force for the development of concepts in all course components. The instructors running the labs were PhD students in physics education who had extensive training in and familiarity with ISLE approach to teaching physics and explicitly incorporated ISLE vocabulary and philosophy into their instruction.

Some preliminary work has been done on the impact of using the ISLE framework to reform only the laboratory component of introductory physics instruction [6]. The purpose of this study was to expand on this work in determining the impact a lab-only adoption of ISLE has on the development of students' scientific abilities. During the course of this study, we found that not all TAs gave accurate feedback to students. This prompted us to investigate the effect that the accuracy of TA feedback had on students' development. This study ended up addressing the following two research questions:

- What impact does a lab-only adoption of ISLE have on the development of students' scientific abilities?
- What effect does the accuracy of feedback that students receive from TAs have on the development of their scientific abilities?

II. COURSE CONTEXT

We selected a second-semester physics course for engineering students to use for this study. While we were given permission to alter the structure of the lab course as we saw necessary without changing the content taught, we were not permitted to make any changes to the associated lecture/recitation course. The course covered the topics of electromagnetism. The labs used in this course are available online [7]. There were ten 3-hour labs in this course.

Twenty-six to 30 students were enrolled in each of the twenty sections of the course. In each section, students were divided into groups of three or four to conduct the experiments and write up their findings.

A. Reformed Labs

The ISLE-based lab activities used in this course consist of a series of hands-on experiments which help students develop their understandings of the course topics while practicing their scientific abilities [8]. Experiments are grouped into three categories: observational, testing, and application experiments. In observational experiments, students develop procedures to explore phenomena and investigate possible relationships such as investigating how the properties of a capacitor affect its ability to store energy. In testing experiments, students design procedures to test hypotheses which were either provided to them or developed as a result of their observational experiments, such as testing the relation for magnetic force by predicting how the force a bar magnet exerts on a scale will change when current is run through a loop of wire positioned between its poles. In application experiments, students use relationships or principles to solve problems or build devices such as an LED flashlight powered by a magnet moving through a coil of wire.

B. Teaching Assistants

Each lab section was led by one Teaching Assistant (TA), supported by one to two Learning Assistants (LAs) [9]. These TAs were undergraduate physics majors (seniors) who had previously completed an equivalent physics course which had not been reformed. As such, the TAs did not have any experience with ISLE prior to this course. The LAs took the course the year before.

TAs were exposed to the ISLE approach to learning physics during a three-hour session before the

beginning of the semester. Additional training sessions were held weekly to prepare the TAs for each week's lab. During these meetings TAs completed the labs and wrote reports as if they were students to better identify challenges their students might face during class. Additional topics such as grading, classroom management, ISLE, and any notes the course coordinator had about specific upcoming labs were addressed during these training sessions as well. Two members of the research team attended these training sessions and shared their findings and suggestions from previous weeks' labs with the Tas and LAs.

C. Scientific Abilities Rubrics

The Scientific Abilities rubrics [10] identify skills and practices which mirror those used by practicing scientists in their work. There are rubrics addressing representational abilities such as creating accurate circuit or force diagrams; skills associated with observational, testing, and application experiments; and abilities which relate to organizing and interpreting data, among others.

Each ability is scored on a scale from 0 (missing) to 3 (adequate), where a 0 represents that there is no evidence of the ability present and a 3 represents that the student or group has demonstrated proficiency in the ability. Scores of 0 and 1 are considered unacceptable while scores of 2 and 3 are considered acceptable. Each rubric contains generalized scoring criteria as can be seen in the sample rubrics found in Figure 1.

In the course described in this paper, each week students submitted lab reports written collaboratively over the course of the lab period. The reports were written during lab time, not at home. These reports served as the main source of data for this study. Each lab explicitly addressed seven scientific abilities from among the abilities listed on the rubrics [11]. The full text of each relevant rubric was included in the lab materials so that students could self-assess while writing their reports. Figure 1 shows two rubrics taken directly from a lab activity.

III. METHODOLOGY

To answer research question #1 we used data from the fall 2017 semester. We scored reports submitted by 50 groups for ten weeks for a total of 500 reports. We did not track TAs as we wanted to have a random sample of groups. In cases where students submitted corrections to their lab reports, only the original reports were scored for research purposes.

While scoring these reports, we noticed that we did not always agree with the scores and feedback given by the TAs. This prompted research question #2. To answer it we used lab reports from fall 2018. That year, we moved to electronic submissions for lab reports and thus had access to more TA feedback. This allowed us to score the reports of ten groups each week taught by each of four different TAs during the fall 2018 semester, for a total of 400 reports, and to compare our scores to those given by the TAs. With these data, we could evaluate the accuracy of the feedback and compare the improvement of the groups taught by each TA to determine if the accuracy of the feedback had any relationship to student development.

While each lab report required students to demonstrate seven abilities, these were not the same seven abilities each week. This means that some abilities were scored only once or twice across the semester. We excluded from our analysis abilities which were scored fewer than three times. As a result, we selected only six abilities for analysis. Table 1 shows the six abilities we analyzed.

To determine whether the improvement in each ability was significant we used independent sample t-tests to compare the first and last time each of the six abilities was scored in order. Regression analysis was used as a secondary means of determining whether the improvement was significant and not due to the effect of outliers. To evaluate the accuracy of each TA's feedback, a modified Cohen's kappa was calculated for each of the four TAs for each week [12]. This value ranges from -1 to 1, with a value of -1 indicating perfect disagreement, a value of 0 representing the amount of agreement expected by random chance, and

Here are the rubric elements your TA will be looking at for this experiment:					
	Scientific Ability	**Missing**	**Inadequate**	**Needs improvement**	**Adequate**
B3	**Is able to decide what physical quantities are to be measured and identify independent and dependent variables**	The physical quantities are irrelevant.	Only some of physical quantities are relevant.	The physical quantities are relevant. However, independent and dependent variables are not identified.	The physical quantities are relevant and independent and dependent variables are identified.
B9	**Is able to devise an explanation for an observed pattern**	No attempt is made to explain the observed pattern.	An explanation is vague, not testable, or contradicts the pattern.	An explanation contradicts previous knowledge or the reasoning is flawed.	A reasonable explanation is made. It is testable and it explains the observed pattern.

FIG 1. Sample rubrics taken from one of the lab activities.

TABLE 1: Abilities analyzed and the changes observed in the percentage of students with acceptable scores

Ability	Description	Percent Change in Acceptable Scores
B2/C2	Can design an experiment to investigate a phenomenon (B2) or test a hypothesis (C2).	0.25 *
B9	Can propose a reasonable explanation for a pattern.	0.05 *
C4	Can make a prediction about the outcome of an experiment based on a hypothesis.	0.05
C8	Can make a judgment about the hypothesis being tested.	0.02
F1	Can communicate the details of an experiment.	0.08
F2	Can communicate the purpose and significance of an experiment.	0.51 **

* $p < 0.05$
** $p < 0.001$

a value of 1 representing perfect agreement. Calculating this value for each lab allowed us to see if the accuracy of their feedback changed over the course of the semester.

To see whether the accuracy of a TA's feedback was correlated with student growth, we plotted the fraction of each TA's students who received scores of 2 or 3 each week. A regression was performed for each TA's students and the slopes were compared to one another. An ANCOVA analysis was performed to determine whether these differences were significant.

IV. FINDINGS

In this section we discuss our findings from the two years of analysis.

A. Changes in Scientific Abilities

Comparing the first and last time an ability was assessed allowed us to determine how students' competency with each ability changed as a result of instruction. As shown in Table 1, all six abilities measured showed improvement when comparing the percentage of students who scored acceptably on the first lab and the percentage of students who scored acceptably on the last lab. However, this improvement was only statistically significant for three abilities: students' abilities to design experiments (B2/C2), students' abilities to propose explanations for their observations (B9), and students' abilities to communicate the purpose and significance of their experiment (F2). Regression analysis of scores over time support these differences being meaningful as opposed to an artifact of sampling.

One possible explanation for why some abilities developed more than others is that the abilities themselves have different difficulties. For example, in the first lab 77% of students demonstrated the ability to communicate the details of their experiments (F1) to an acceptable level while only 31% of students demonstrated the ability to communicate the significance of their results (F2). This means that some abilities had much more room for growth than others.

We also observed that, despite the general growth of student abilities, student scores fluctuated week to week more than they had in previous studies. One possible explanation for this is that previous studies were conducted in mechanics courses, where the difficulty of the physics content explored in each lab activity does not vary as much as it does in an electricity and magnetism course.

B. Accuracy of Feedback

Figure 2 shows the modified Cohen's kappa value for each of the four TAs studied across all 10 weeks of the course (two TAs did not report scores for the tenth lab and so their kappa values could only be calculated to week 9). We found that the accuracy of the feedback provided by all TAs improved over the course of the semester. For the first few weeks, TAs were scoring much too high, as can be seen in Figure 3 which shows the number of each score given by TA 1 and the researchers for Lab 1 and Lab 10. In this figure, the gray boxes on the diagonals represent where the researcher and TA assigned the same score to an ability. The boxes beneath the diagonal represent where the TA

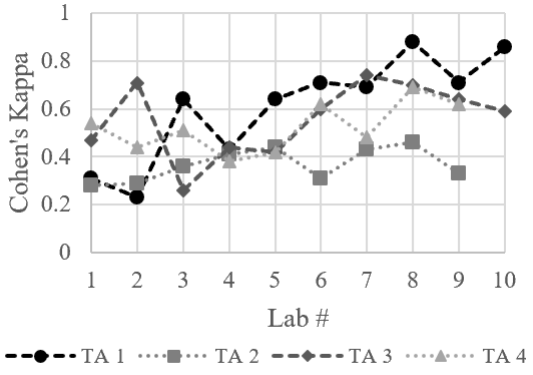

FIG 2. Agreement between TAs and researchers.

scored lower than the researcher. It is encouraging to see that feedback accuracy improved for all TAs over time. The average kappa value across all labs was calculated to estimate the overall accuracy of a TA's feedback. From this we found that TA 1 gave relatively accurate feedback (0.61), TA 2 gave relatively inaccurate feedback (0.37), and TAs 3 and 4 were somewhere in the middle (0.56, 0.52).

Figure 4 shows the fraction of student scores which were rated as acceptable (2 or 3) each week. From this graph we see that students of TA 1, who provided the most accurate feedback, showed the largest improvement and students of TA 2, who provided the least accurate feedback, showed the least improvement. ANCOVA analysis of these data support that this difference is significant ($p = 0.037$).

Lab 1

TA Score \ Researcher Score	0	1	2	3
0	4			
1	2	6	2	
2		2	2	
3	4	14	19	15

Lab 10

TA Score \ Researcher Score	0	1	2	3
0	1			
1	5	17	1	
2			18	
3		2		26

FIG 3. TA and researcher scoring for Lab 1 and Lab 10.

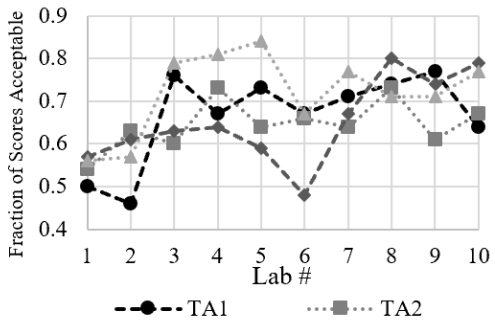

FIG 4. Fraction of the rubric scores which were scored as acceptable (2 or 3) for each TA.

V. SUMMARY AND LIMITATIONS

In this paper we sought to determine whether the previously reported benefits of the ISLE-based labs could be replicated within the context of a less-ideal, more traditional course. Making no changes to the associated lecture course and having minimally-trained undergraduate TAs run the lab sections, we still found that using ISLE-based labs produced small but significant improvement in students' scientific abilities over the course of one semester. We also found that the accuracy of the feedback students received from these TAs correlated with student improvement.

This suggests that students can benefit just from altering the lab component of a general physics course without making more substantial changes to the rest of the course. While we would still advocate for as much reform as possible, we recognize that completely altering courses may sometimes be impractical. Even if larger reforms are impossible, simply changing the lab activities does have a beneficial effect for students.

Our findings also suggest that any reform efforts using the Scientific Abilities rubrics must include training for the instructors on how to accurately use the rubrics to score student work. Just as we began each semester with an intensive training session on ISLE philosophy, similar upfront training with the rubrics may be necessary to reduce the time it takes for instructors to start providing accurate feedback using the rubrics.

The study has several limitations. In our analysis we did not consider the roles the LAs played in the classroom or the impact they had on student development. We also did not conduct classroom observations during the Fall 2018 semester, so we cannot rule out the possibility that the different rates of improvement between each TA's students can be attributed to the TA's behavior in the classroom or other unmeasured differences between TAs.

Our next steps will focus on improving the training TAs receive in order to use the Scientific Abilities rubrics to provide feedback more accurately.

[1] A. Hofstein and V. N. Lunesta, "The laboratory in science education: Foundations for the twenty-first century", *Sci. Educ.* **88** 1, 28-54 (2004).

[2] N. G. Holmes and C. E. Wieman, "Introductory physics labs: We can do better" Phys. Today **71** 1 38-45, (2018).

[3] E. Etkina and A. Van Heuvelen, "Investigative Science Learning Environment – A science process approach to learning physics," Research-based Reform of University Phys. **1** 1, (2007).

[4] E. Etkina, A. Van Heuvelen, D. Brooks, and D. Mills, "Role of experiments in physics instruction: A process approach", Phys. Teach. **40**, 351 (2002).

[5] E. Etkina, A. Karelina, M. Ruibal-Villasenor, M. Rosengrant, D. Jordan, and C. E. Hmelo-Silver, "Design and reflection help students develop scientific abilities: Learning in introductory

physics laboratories", J. of Learning Sci. **19** 1, 54-98 (2010).
[6] D. Demaree and Y. Lin, "Assessing ISLE labs as an enhancement to traditional large-lecture courses at the Ohio State University", in 2005 Physics Education Re-search Conference (Salt Lake City, UT, August 2005), edited by P. Heron. L. McCul-lough, and J. Marx, AIP Conf. Proc. 818, 101 (2006).
[7] Lab manual can be found under 2017 Rutgers – Electromagnetism (for engineering majors) at https://sites.google.com/site/scientificabilities/ISLE-labs.
[8] E. Etkina, S. Murthy, and X. Zou, "Using introductory labs to engage students in experimental design", Am. J. Phys. **74**, 979 (2006).
[9] V. Otero, S. Pollock, and N. Finkelstein, "A physics department's role in preparing physics teachers: The Colorado learning assistant model", Am. J. Phys. **78** 11, 1218-1224, (2010).
[10] E. Etkina, A. Van Heuvelen, S. White-Brahmia, D. Brookes, M. Gentile, S. Murthy, D. Rosengrant, and A. Warren, "Scientific abilities and their assessment", Phys. Rev. ST Phys. Educ. Res. **2**, 020103 (2006).
[11] Rubrics can be found at https://sites.google.com/site/scientificabilities/rubrics
[12] J. Cohen, "Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit", Psych. Bulletin **70** 4, 2013-220, (1968).