PHYSICAL REVIEW PHYSICS EDUCATION RESEARCH **19,** 020158 (2023)

Comparing students' learning and development of scientific abilities with apparatus-based versus video-based experimentation

David T. Brookes, Mc Kenna Wallace, and Michael Nelson
*Department of Physics, California State University, Chico, 400 W. 1st Street,
Chico, California 95929-0202, USA*

Anna Karelina
Saint Mary's College of California, 1928 St Mary's Road, Moraga, California 94575, USA

Peter Bohacek
Two Rivers High School, 1897 Delaware Avenue, Mendota Heights, Minnesota 55118, USA

Matthew Vonk
*Department of Physics, University of Wisconsin, River Falls, 410 S 3rd Street,
River Falls, Wisconsin 54022, USA*

Eugenia Ektina
*Graduate School of Education, Rutgers University, 10 Seminary Place,
New Brunswick, New Jersey 08901, USA*

(Received 28 April 2023; accepted 18 September 2023; published 16 November 2023)

[This paper is part of the Focused Collection on Instructional labs: Improving traditions and new directions.] In this paper, we describe the results of a research project whose goals were to (1) develop and implement video-based experimental investigations using the Investigative Science Learning Environment (ISLE) approach and (2) study how students who engage with video experiments develop scientific abilities and learn physics ideas in comparison to students who do the same investigations using physical apparatus. We developed six parallel ISLE-based investigations for the students to engage in, either with apparatus or with video arrays created in the Pivot platform. We found that substituting 30% of the apparatus-based activities with video-based activities did not affect student development of conceptual physics knowledge. On the other hand, the development of certain scientific abilities was significantly affected by whether students experimented with physical apparatus or used video experiments.

DOI: 10.1103/PhysRevPhysEducRes.19.020158

I. INTRODUCTION

In this paper, we describe the results of a research project whose goals were to (1) develop and implement video-based experimental investigations using the Investigative Science Learning Environment (ISLE) approach [1] and (2) study how students who engage with video-based experiments develop scientific abilities and learn physics ideas in comparison to students who do similar investigations using physical apparatus.

A. Background

Research in the use of computer-based technology to help students learn physics is extensive, stretching back to microcomputer-based-laboratories [2]. We cannot hope to do it justice in this paper and will only provide a few touchstone examples. Studies have investigated comparisons between students working with real equipment and physics simulations. Finkelstein *et al.* showed that students who learned dc circuits through PhET simulations were better able to build and analyze real circuits than those who learned dc circuits with physical apparatus [3]. Stelzer *et al.* showed that students learn physics concepts better and retain their knowledge longer from a well-designed video presentation (designed according to the principles of multimedia learning [4]) compared to students who read a text of the presentation or students who read the textbook [5]. Without delving too much further into the topic, there is ample

---

evidence that computer-based technologies with simulations and videos can help students both learn normative physics ideas *and* some practical experimental skills and knowledge.

B. The educational value of physics laboratories

To understand the motivation of our present study, we need to narrow our focus and ask: What is the pedagogical value of having students engage in experimentation in the physics laboratory? Holmes *et al.* showed that students who engaged in traditional labs did not receive any advantage in the development of normative physics knowledge compared to students who did not do labs at all [6]. Complementary to that research, Etkina *et al.* compared the exam performance of two groups of students: Those who engaged in traditional "cookbook" labs where they were told what to do and what to measure versus students who were guided to design an experiment to achieve certain epistemic goals but were not told what to do or what to measure ("design labs"). They found that students who engaged in design labs spent a great portion of their time struggling with the experiments and reflecting on the process of construction of knowledge through special reflection activities. Despite all the time spent on design and reflection, they did as well on exam problems as the students who performed "cookbook" experiments and instead of reflection, spent the time that was saved solving additional physics problems [7].

One possible benefit of physics lab lies in the growth of students' abilities to think and reason like a physicist. These scientific abilities are fully explored in Ref. [8] and explained in more detail in Sec. II B below. Over the years Etkina *et al.* conducted a set of studies of the labs that were an integral component of courses that used the ISLE approach. They found that students who engaged in experimental design, structured around scientific abilities rubrics [8], learned to engage in scientific practices. (An example of such a design lab is described in Sec. II B.) By the end of one semester, about 80% of them were able to distinguish between hypotheses and predictions, identify and evaluate experimental uncertainty, identify relevant assumptions in predicting the outcome of testing experiments, and were able to compare the results of two independent experiments aimed to determine the same physical quantity [9]. Students were able to apply these abilities significantly better when given a transfer task than students who did traditional lab experiments guided by precise instructions about what to do [7].

C. The affordances and constraints of video-based experimentation

As the focus of the ISLE approach is on students *doing* physics as they are learning normative concepts, our goal is to investigate the affordances and constraints of student participation in authentic scientific practices when they work with video experiments as opposed to real equipment in an ISLE setting. The affordances of video-based experimentation could provide many advantages over experimentation with physical equipment:

1. If funding is scarce, students can do experiments without expensive equipment and without being physically present in a lab.
2. Video experiments present observational data with high accuracy and precision. This can reduce the cognitive load for students trying to identify patterns in those data. In addition, all groups work with the same experimental setup, and this setup has no issues (which might happen with physical apparatus). Therefore, students can spend more time figuring out a functional relation between variables or pondering the assumptions that are implicit in a model (we found that students struggle the most with assumptions in Ref. [9].)
3. Video-based experiments allow students to play them frame by frame, rewind to see if they missed anything, and work with phenomena that occur too quickly in real life [10].

However, videos may come with certain constraints:

1. If the video-based labs lack tools for students to work collaboratively in the online environment, students may end up working in greater isolation and lose some of the benefits of social interactions [11].
2. Students are not able to manipulate the equipment, which was found to be an important aspect of human cognition [12,13].
3. There is a possibility that videos may reduce or remove some of the variability of data collected from physical apparatus. Consequently, students may have fewer opportunities to learn about random experimental uncertainties. Video-based lab developers may need to pay special attention to the design of such labs to make sure that experimental noise is built into the environment.

D. Research questions

Narrowing our view to the value of physics labs in terms of students' development of scientific abilities, all the studies described above were conducted in labs where students manipulated physical apparatus. While research shows that using video experiments helps students learn how to develop and "break" models [14], there are no studies comparing different aspects of students' development of scientific abilities when they design their own experiments using physical apparatus versus working with videoed experiments. The work reported in this paper is the first attempt to investigate the differences in student learning with physical equipment or with videoed experiments in design labs that are integrated into a holistic learning environment [15].

In our current research project, we wanted to know: Is it possible to replicate student development of scientific abilities using video experiments instead of ones using apparatus? If we replace physical equipment with video experiments, is it possible to recreate similar cognitive conditions (where students make decisions about what experiment to conduct and what data to collect, etc.) as in our prior research; conditions that have been shown to lead to the development of scientific abilities? And if it is possible, how does this replacement affect students' development of scientific abilities and student learning of normative physics ideas? While these were purely research questions before the beginning of the COVID pandemic, now these are questions that every teacher may be asking. How effective is student learning when they cannot manipulate physical apparatus? What are students gaining? What are they losing?

We set out to answer the following two research questions:

1. What are the differences in how students develop scientific abilities when experimenting with videos versus apparatus in ISLE-based labs?
2. How does experimenting with video affect student learning of normative physics concepts in an ISLE-based course such as (for example) Newton's second and third laws as compared to students who learn the same physics ideas by experimenting with apparatus?

Below, we describe in detail the epistemic structure of the ISLE approach and the role of experimentation in this approach [16,17]. The epistemic structure and the guided inquiry of the experimentation process are, we believe, how the ISLE approach fosters the growth of students' abilities to reason like a scientist. This will illustrate the challenge of how to replace an apparatus-based lab with videos while still allowing students to make the kinds of decisions about what to measure, what data to collect, and how to analyze it. These features are the hallmark of the ISLE approach and how it fosters scientific reasoning abilities.

II. THE ISLE APPROACH TO LEARNING PHYSICS

A. Summary

One of the major goals of the ISLE approach is that students learn physics by engaging in processes that mirror the activities of physicists constructing and applying knowledge [18]. Based on the history of physics and observations of practicing physicists, the creators of the ISLE approach identified those activities and they became an integral part of students' learning. In ISLE, all experiments that students encounter in a course are grouped into three big categories: observational experiments (students collect, analyze data, and propose models or relations, hypotheses), testing experiments (students use proposed models, relations, or hypotheses to make predictions about their outcomes), and application experiments (students combine several tested models or relations to solve a practical problem or determine an unknown physical quantity). While engaging in epistemologically authentic scientific investigations [19] centered on the experimental activities described above, students develop scientific habits of mind or scientific abilities [8]. To develop the normative concepts of physics as their own ideas, ISLE students repeatedly go through the steps that mimic scientific exploration. They:

1. Observe and describe physical phenomena in carefully chosen, simple observational experiments and identify patterns in collected data. Or students can design their own experiments to investigate a specific phenomenon,
2. Develop multiple models or explanations for these patterns,
3. Design new testing experiments whose outcomes they can predict using these models or explanations,
4. Compare the outcomes of the testing experiments with the predictions,
5. Revise the models or explanation if necessary, examine assumptions, or go back to collect more observational data,
6. Design application experiments to apply tested and not-ruled-out models or explanations for practical purposes (build devices, determine the values of physical quantities, etc.)
7. Work in groups of 3–4, use small whiteboards, and then share their findings in a whole class discussion,
8. Use representations other than algebra or calculus to analyze data and make predictions.

After they have explored and analyzed the phenomenon, students read the textbook and compare the ideas that they have constructed on their own with the ideas presented in the book. The combination of these features applies to every conceptual unit in the ISLE approach to learning physics (such as Newton's laws or circular motion, etc.) [18].

B. The role of experimentation in the ISLE approach

In the ISLE approach, the three types of experiments—observational, testing, and application, serve as the backbone of student learning. These three types of experiments [16] and the epistemic role that they play is one of the features that distinguish the ISLE approach from other curricular approaches that emphasize the process of doing physics. ISLE-based labs are not an add-on to the "lecture course" but an integral part of how students learn physics through the ISLE approach. Therefore, these labs are not only different from traditional labs in which students follow direct instructions but from other types of reformed labs such as structured quantitative inquiry labs [20] and Maryland Scientific community labs [21]. Research on ISLE-approach labs can be found in many publications [7,9,22–26]. Laboratories that follow the ISLE approach typically share the following features:

TESTING EXPERIMENT: DIRECTION OF THE FORCE EXERTED BY THE MAGNETIC FIELD ON A WIRE

Available equipment: A horseshoe magnet whose poles are known (Red: North, White: South), a scale, an assortment of rigid wires (mounted on plastic backings. They look a bit like slides.), a voltage source, multimeter, connecting wires, ring stand, meter stick, PASCO current balance.

Think of how you can use this equipment to test the right-hand rule for the direction of the force exerted by the magnetic field on a current carrying wire. Hint: What physical quantities can you determine using the scale?

a. First, recall the right-hand rule for the magnetic force exerted by the magnetic field on a current carrying wire. Write what quantities it relates and describe the rule with a picture and using words. Consider the available equipment and how you could use it to achieve the goal of the experiment. Brainstorm and write down your potential experiments. Think ahead about what you will measure and how you will measure it.
b. Describe the experimental procedures you have chosen. The description should contain a labeled sketch of your experimental setup, an outline of what you plan to do, what you will measure, and how you will measure it. Explain in detail how you will experimentally measure the direction of the force exerted by the magnetic field on the wire. Hint: The reasoning here is more complicated than it seems at first. Use force diagram(s) and Newton's second and third laws to help.
c. Use the hypothesis you are testing to make qualitative predictions for the reading of the scale (more than some value, less than some value) for every experiment that you plan to run. Show the reasoning used to make the prediction with force diagrams. Call your TA over once you have done this but before you turn on the current.
d. Perform the experiments and record the outcomes.
e. Did the outcomes match your prediction? If not, list possible reasons.
f. Based on your prediction and the experimental outcome, make a judgment about the right-hand rule you developed previously.

RUBRICS

RUBRIC A: Ability to represent information in multiple ways				
Scientific Ability	Missing	Inadequate	Needs some improvement	Adequate
Is able to construct a force diagram. (A5)	No representation is constructed.	Force diagram is constructed but contains major errors such as the forces not relating to any interactions, mislabeled or not labeled force vectors, length of vectors, wrong direction, extra incorrect vectors are added, or vectors are missing.	Force diagram contains no errors in vectors but lacks a key feature such as labels of forces with two subscripts or vectors are not drawn from single point or axes are missing.	The diagram contains no errors and each force is labeled so that it is clearly understood what each force represents.
RUBRIC C: Ability to design and conduct a testing experiment (testing an idea/hypothesis/explanation or mathematical relation)				
Is able to distinguish between a hypothesis and a prediction. (C3)	No prediction is made. The experiment is not treated as a testing experiment.	A prediction is made but it is identical to the hypothesis.	A prediction is made and is distinct from the hypothesis but does not describe the outcome of the designed experiment.	A prediction is made, is distinct from the hypothesis, and describes the outcome of the designed experiment
Is able to make a reasonable prediction based on a hypothesis. (C4)	No attempt to make a prediction is made.	A prediction is made that is distinct from the hypothesis but is not based on it.	A prediction is made that follows from the hypothesis but does not incorporate assumptions.	A correct prediction is made that follows from the hypothesis and incorporates assumptions.
Is able to make a reasonable judgment about the hypothesis. (C8)	No judgment is made about the hypothesis.	A judgment is made but is not consistent with the outcome of the experiment.	A judgment is made and is consistent with the outcome of the experiment but assumptions are not taken into account.	A reasonable judgment is made and assumptions are taken into account.

FIG. 1. Example of an ISLE lab testing experiment.

1. In the ISLE approach, labs are not a separate course but an integral part of physics learning. It is in the labs that students encounter new phenomena which they learn to explain in other parts of the course and later test those explanations back in the labs. All students in the course do the same lab during the same week.
2. All experiments that students design and perform are grouped into three big categories according to their epistemic goals, as described above [16].
3. In ISLE-approach labs, students design their own experiments being guided by questions that prompt them what to think about, not what to do [22].
4. When designing, performing, and describing the results of the experiments, the students use scientific abilities rubrics [8] to guide and improve their work. The instructors use the same rubrics to provide feedback to the students and grade their submissions.

In Fig. 1, we show an example of one of the experiments that the students do in ISLE-based labs. Here we describe how scientific abilities rubrics are used in that lab to help students write their lab reports and how the same rubrics help instructors assess student work. The lab is done after students have completed an activity in class constructing a hand rule for the direction of the magnetic force exerted on a current-carrying wire. In the lab, they need to test the new rule by (a) designing experiments, (b) making predictions of experimental outcomes using the hypothesis under test, and (c) forming a conclusion based on the outcome of the experiment. To help them design, conduct, and report on the experiments, they are provided with self-assessment scientific abilities rubrics that contain descriptors of different levels of achievement [8]. There are seven rubrics that help students in all activities. Those are multiple representations rubric, observational experiment rubric, testing experiment rubric, application experiment rubric, data collection and analysis rubric, evaluation rubric, and communication rubric. Each rubric has multiple rows that describe different aspects of the assessed ability. We call these rows "rubric items." There are all together over 40 rubric items that help students with various aspects of different types of experiments and multiple representations, but for one lab experiment, we usually use only 3–5 of those rubric items not to overwhelm students and to focus on the most important aspects of the particular investigation. These rubric items are embedded in the handout that students receive (see Fig. 1).

The reader can see from the example in Fig. 1 that students need to design their own experiments, conduct an analysis that will allow them to make the predictions, run the experiment, collect and analyze data, and develop their own judgment. The chosen rubric items help them figure out how to use the scale to infer the value of the magnetic force and how to reason about collected data. The reader can see from the numbers of the rubrics items that there are more items in the multiple representations rubric but they are not relevant as the students do not use motion diagrams, ray diagrams, and other representations mentioned in rubric A. In the testing experiment rubric C, there are also more items but they are not relevant to the task that students are engaging in. It is also important to understand that scientific abilities rubrics can be used both as a formative assessment tool, scaffolding students' writing and providing them with feedback, and as a research tool to assess students' development of scientific abilities as described in Ref. [9].

III. DESCRIPTION OF THE STUDY

A. ISLE-approach labs using apparatus or videos

In 2017–2018, the team of authors developed six ISLE-approach experimental sequences using a cloud-based platform called Pivot Interactives described in Ref. [27]. Pivot Interactives contains arrays of videos, where all the videos in a set show the same scenario occurring, but each video has a unique set of experimental parameters (like mass, angular speed, spring constant, trial number, etc.). This enables students to select the experimental conditions they would like to investigate. Supporting questions in the Pivot platform help students analyze and interpret the data that they collect from the videos. The platform has interactive measurement tools that allow students to make decisions about how, what, and when to measure. These choices make video experimentation in Pivot as close as possible to ISLE-approach experimentation with apparatus where students have to engage in experimental design as elaborated in Sec. II B. While in the video-based labs, the students cannot design an original experiment, they can vary the physical quantities involved (as if they are collecting data in observational experiments), predict experimental outcomes (as if they are conducting testing experiments), and compare the results of two independent experiments when measuring a physical quantity (as if they are doing application experiments). The experimental sequences (observational, testing, and application experiments) developed for this project are Newton's second law, Newton's third law, circular motion, momentum conservation, vibrational motion, and wave motion. These sequences are accompanied by supporting questions based on the ISLE approach. We called these sequences vISLEs, for video-based ISLEs, and are available for free at Ref. [28].

B. Experiment design, setting, and participants

In the present study, we investigated student learning in a medium-size (about 90 students) algebra-based introductory mechanics course at a public four-year university in Northern California. The course followed the ISLE approach in all aspects and lab activities were fully integrated into the course. Our study consisted of matched apparatus-based and video-based activities used in five lab sessions. We will call the labs where students use ISLE-based activities involving physical equipment, "apparatus-based labs" and the labs where the students use vISLE activities, "video-based labs."

The experiment was repeated over two semesters: Fall 2018 and Spring 2019. In Fall 2018, there were 86 students in the course and there were 93 in Spring 2019. It is important to note that the university is located in the area where fires are common and in Fall 2018, many families of the students in the course were affected by a large fire that destroyed a nearby town.

Each week there were two 75-min large-room meetings where all students in the course worked collaboratively in groups of 4 using a whiteboard while the instructor and learning assistants circulated around the room, helping students. In addition to large-room meetings, the students attended 170-min lab meetings with approximately 24 students in each lab section. Lab sections (total of 4) met in between the large-room meetings. The groups of four students were kept the same throughout the semester. There were 15 lab meetings during the semester. Five lab meetings implemented our experimental intervention of equivalent experiments using either apparatus or video. The other ten labs used apparatus as usual and were the same for all four lab sections. One section (section 1) only did apparatus-based labs (AAAAA), and one section (section 4) did video-based labs (VVVVV) during the 5 weeks of the experiment. The other two sections (2 and 3) represented a counterbalanced condition of two apparatus-based followed by three video-based labs (AAVVV) (section 2), or two video-based ones followed by three apparatus-based labs (VVAAA) (section 3). The assignment of the sections

TABLE I. Summary of intervention for both Fall 2018 and Spring 2019.

	Section 1	Section 2	Section 3	Section 4
Week 5: Newton's second and third laws	Apparatus	Apparatus	Video	Video
Week 7: Circular motion	Apparatus	Apparatus	Video	Video
Week 9: Momentum	Apparatus	Video	Apparatus	Video
Week 13: Vibrational motion	Apparatus	Video	Apparatus	Video
Week 15: Mechanical waves	Apparatus	Video	Apparatus	Video

to the different conditions was random; we renumbered them to a sequential order in reporting our results to lessen the reader's cognitive load. Lab assignments for different groups and the schedule of labs are shown in Tables I and II. Note that all but one section was taught by the same instructor, the leader of this project. One section (No. 1, in which students only did apparatus labs in the Fall of 2018) was taught by a different instructor who was also skilled in the ISLE approach.

In each lab, there were two or three different experiments; each of them was framed in a way that the students could engage in the same activities and thought processes whether they had physical apparatus or arrays of videos. The wording of the student handout in both versions was identical except for references to the video. The only exception was the circular motion lab, where it was not possible to create an equivalent quantitative observational experiment for the apparatus-based version. The apparatus condition did a different activity for the observational experiment compared with the video-based lab. See Ref. [39] for a comparison of an apparatus-based lab and equivalent video-based lab.

The students in the apparatus-based labs needed to create a lab report that address the guiding questions in their handouts while the students in the video-based labs entered their responses to the same guiding questions into text boxes. The students worked in groups of 4 and submitted their reports in pairs. The lab reports for apparatus-based labs were submitted via Google Drive and the lab reports for the video labs were submitted via the Pivot Interactives system. The instructor provided feedback to both groups in the margins of the submitted documents.

IV. DATA COLLECTION AND ANALYSIS

A. Summary

We collected the following data for this study:

- Student lab reports for two semesters, see Table III (a total of 242 group lab reports).
- Student exam work (total of 86 in the Fall of 2018 and 93 in the Spring of 2019 individual exams).

B. Lab reports

We collected lab reports from multiple lab activities across all sections and from both semesters of the experiment. For research purposes, gathered lab reports were scored using scientific abilities rubric items appropriate to each experiment [8]. Two scorers worked on scoring lab reports. One person was an expert, who coauthored numerous papers describing the use of rubrics for research and the other one was a specially trained scorer for this project. After the second scorer was trained in the use of the rubrics, they scored together about 10% of the reports for each experiment. They compared the scores and discussed the discrepancies, finding the reason for the discrepancy and arriving at a common score after discussion. Then each scored the rest of the lab reports and calculated the interrater reliability of Cohen's kappa which were all above 0.6 (for some rubrics > 0.8). This indicates moderate to substantial agreement. Lab report scores were then

TABLE II. Schedule of experimental intervention: ISLE-based apparatus and vISLE activities used in the study.

Week 5: Newton's second laws	Experiment 1: Observational experiments: Exploring the relationship between acceleration, mass, and force [29].
	Experiment 2: Observe and find a pattern (two scales) [30].
	Experiment 3: Testing Experiments: Interactions between different objects during collisions [31].
Week 7: Circular motion	Experiment 1: Observational experiment: Exploring circular motion [32].
	Experiment 2: Testing experiment: Circular motion [33].
Week 9: Momentum	Experiment 1: Devise a new physical quantity. Observational experiments [34].
	Experiment 2: Testing whether the physical quantity is constant for a new isolated system [35].
Week 13: Vibrational motion	Experiment 1: Testing experiment: Testing the mathematical relation you derived for ω [36].
	Experiment 2: Application experiment: Measure the spring constant of a spring two different ways [37].
Week 15: Mechanical waves	Experiment 1: Observational Experiment: Exploring Properties of Transverse Mechanical Waves [38].

TABLE III. Summary of lab reports data gathered across 2 semesters.

	Observational experiment	No.	Testing experiment	No.	Application experiment	No.	Total no.
Apparatus	Fall 2018, waves	21	Fall 2018, circular motion	23	Fall 2018, vibrations	22	
	Spring 2019, waves	22	Fall 2018, vibrations	22	Spring 2019 vibrations	11	
	Total	43	Total	45	Total	33	121
vISLE	Fall 2018, circular motion	22	Fall 2018, vibrations	18	Fall 2018, vibrations	18	
	Fall 2018, waves	19			Spring 2019 vibrations	22	
	Spring 2019, waves	22					
	Total	63	Total	18	Total	40	121

combined according to the type of experiment (observational, testing, and application). These combined scores were used to compare students' development of scientific abilities between apparatus-based and video-based labs.

C. Exam questions

Across the two semesters, we embedded specific exam questions that were designed to assess both students' conceptual understanding of ideas related to topics they encountered in lab activities that were part of the experimental intervention; and their scientific reasoning abilities. In particular, in both Fall and Spring semesters, we embedded

1. A question unrelated to the topics of the intervention (an energy question) focused entirely on representational knowledge in the context of energy. This question targeted scientific abilities and physics topics unrelated to the types of scientific abilities and topics involved in the intervention labs. This provided a covariate for the data analysis if it turned out that there were large discrepancies between the

TABLE IV. Rubric for research-based scoring of student responses to the pendulum question. Note that the student had to explicitly meet the criteria described in the right-most box to achieve a maximum score on that particular rubric item.

0	1	2	3
		Data analysis	
No analysis done or does inappropriate things like averaging across trials for different degrees.	Finds 6 distinct averages but doesn't find uncertainties in each trial. Or makes a T_{avg} versus θ plot without error bars.	Finds 6 averages and calculates uncertainties but uncertainties are calculated incorrectly (e.g., assumed instrumental rather than random spread)	Finds averages and correctly calculates uncertainties properly and/or plots with error bars. Full credit is earned if the spread is halved.
		Judgment	
No judgment. OR, ONLY refers to period formula rather than data in making judgment OR, inconsistent judgment where conclusion directly contradicts their calculated confidence interval(s) or uses fake evidence.	Consistent judgment—concludes yes/no there is/is not a trend, consistent with calculated uncertainties without actually talking explicitly about those uncertainties, or talks about them in a super vague, hand-wavey kind of way. OR, concludes there IS a trend without any uncertainties calculated.	Consistent judgment—concludes yes/no there is/ is not a trend with reference to the confidence interval, but does so in a selective way, e.g., looking at 10° versus 20° while ignoring 40°, 50°, and 60° or vice versa.	Consistent judgment—concludes yes/no there is/ is not a trend with explicit reference to confidence intervals across the entire dataset.
		Communication	
No explanation	Explanation takes effort to comprehend or is clear but is not consistent with the data analysis.	Explanation is easy to comprehend: Explain how they calculated uncertainties. Even if the judgment is incorrect, it is clear how they arrived at it and it is consistent with the data analysis.	

TABLE V. Rubric for research-based scoring of student responses to the energy question. Note: The student had to be explicit to receive a score of 2. To be explicit the answer had to include details indicating they understand if there is K and U_g, then system objects must still be moving and have not reached the zero gravitational potential energy.

0	1	2
	System and objects	
No mention of objects and/or system. Or objects/system are inconsistent w/ bar chart.	Objects and system are identified, but with minor inconsistencies (e.g., earth omitted when talking about Ug).	Objects and the system identified (what objects are in the system and what objects are not in the system). Earth explicitly included in the system.
	States	
No mention of initial and final states or both states are vaguely and/or inconsistently identified.	Identifies 1 state clearly and consistently, but not the other is vague and/or inconsistent.	Identifies both states clearly and consistently. Each object's position and movement must be clearly specified.
	Process	
No process mentioned or completely inconsistent with the bar chart.	Process partially consistent, some changes accounted for, but some omitted. (E.g., poor accounting for ΔU_{int}).	Process consistent with the bar chart, all changes accounted for in the description.

different lab sections. The questions used in Fall 2018 (Fig. 7) and Spring 2019 (Fig. 8) are in the Appendix.

2. A question that examined students' ability to explain Newton's second and third laws and make a clear distinction between the role of the two in the Newtonian model of the physical world. The two questions (Figs. 9 and 10) are in the Appendix.
3. A question that presented students with a set of data of pendulum periods for different angles of swing. This question required students to analyze a dataset, calculate experimental uncertainties, and draw a conclusion from the data. The question (Fig. 11) was identical in both semesters and can be found in the Appendix.

All of these questions were scored on rubrics designed by the research team. We designed the rubrics based on a survey of students' responses, discussing what sorts of thematic elements we observed and what constituted an adequate answer to the question. The rubrics for the pendulum question and for the energy question are shown in Tables IV and V. The rubrics were used for research purposes, not for student feedback.

V. FINDINGS

A. Lab reports

Below, we present the findings from scoring students' lab reports using scientific abilities rubrics [40] (see

TABLE VI. Rubric items used for the analysis of observational experiment reports.

Ability	Missing (0)	Inadequate (1)	Needs improvement (2)	Adequate (3)
Is able to identify a pattern in the data (B7)	No attempt is made to search for a pattern	The pattern described is irrelevant or inconsistent with the data	The pattern has minor errors or omissions. Terms proportional are used without clarity: is the proportionality linear, quadratic, etc.	The pattern represents the relevant trend in the data. When possible, the trend is described in words.
Is able to represent data graphically (A11)	No graph is present	A graph is present but the axes are not labeled. There is no scale on the axes. The data points are connected.	The graph is present and axes are labeled but the axes do not correspond to the independent and dependent variable or the scale is not accurate. The data points are not connected but there is no trendline.	The graph has correctly labeled axes, independent variable is along the horizontal axis, and the scale is accurate. The trendline is correct.

Scientific ability	p	**Who did better?**	**Statistical test**	N	**Number of labs analyzed**
Rubric B: Ability to design and conduct an observational experiment					
Ability to identify patterns in data (B7)	0.0003	Apparatus	F-H[a]	67	2 (Waves, fall and spring.)
Quality of graphical plots (A11)	< 0.0001	Video	F-H[a]	46	1 (Newton's second law, fall.)
Rubric C: Ability to design and conduct a testing experiment					
Design a testing experiment (C2)	0.45	Not significant	F-H[a]	40	1 (Vibrations testing experiment, fall)
Make a prediction based on a hypothesis (C4)	0.04	Unclear[b]	F-H[a]	40	1 (Vibrations testing experiment, fall)
Identify assumptions (C5)	0.0006	Video	F-H[a]	40	1 (Vibrations testing experiment, fall)
Decide if outcome matches prediction (C7)	0.07	Not significant	F-H[a]	40	1 (Vibrations testing experiment, fall)
Make a judgment about hypothesis (C8)	0.02	Unclear[b]	F-H[a]	40	1 (Vibrations testing experiment, fall)
Rubric D: Ability to design and conduct an application experiment					
Experimental design (D2)	0.27	Not significant	F-H[a]	73	2 (Spring constant, fall & spring)
Make a judgment (D4)	0.38	Not significant	F-H[a]	73	2 (Spring constant, fall & spring)
Evaluate results w/ second method (D5)	0.02	Video	F-H[a]	73	2 (Spring constant, fall & spring)
Use mathematical procedure (D7)	0.04	Unclear[b]	F-H[a]	73	2 (Spring constant, fall & spring)
Identify assumptions (D8)	0.0003	Video	F-H[a]	73	2 (Spring constant, fall & spring)
Rubric G: Ability to analyze data					
Identify sources of uncertainty (G1)	0.007	Apparatus	Chi sq.	242	6 labs combined data, fall & spring
Evaluate uncertainties (G2)	0.0002	Apparatus	F-H[a]	242	6 labs combined data, fall & spring
Minimize uncertainties (G3)	< 0.0001	Apparatus	F-H[a]	242	6 labs combined data, fall & spring
Record data appropriately (G4)	0.26	Not significant	Chi sq.	242	6 labs combined data, fall & spring
Analyze data appropriately (G5)	< 0.0001	Apparatus	Chi sq.	242	6 labs combined data, fall & spring

[a] The Freeman-Halton extension of the Fisher exact test for a 2×4 contingency table is necessary when cells have less than 5 data points, violating one of the key assumptions of the chi-squared test.

[b] Some results are statistically significant on a chi-squared or Freeman-Halton test. However, these tests test whether two distributions are different from each other, *not* whether two averages are different from each other. In these cases the score distributions were significantly different, but the average scores were fairly close.

FIG. 2. Results and comparisons across conditions of scoring scientific abilities rubrics to compare two conditions.

Table III for the types of experiments used in the study). The results of the comparisons of different conditions (apparatus or video) are reported in Fig. 2. We follow with a detailed analysis of rubric scores. The analysis will help the reader better understand comparisons in Fig. 2. Our data do not reflect the changes over a time period but represent cumulative averages for all lab pairs that participated in the study.

1. Observational experiments

We used the following rubrics for the observational experiments and data analysis (see Table VI).

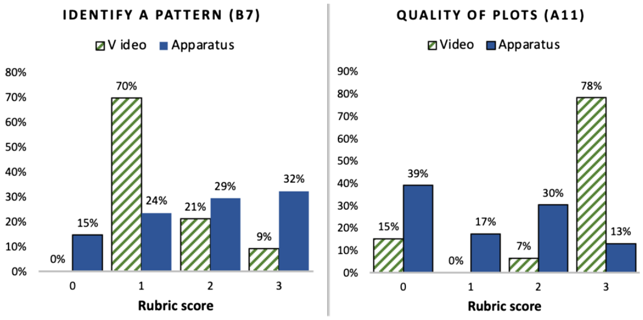


FIG. 3. Comparing abilities involved in designing and conducting an observational experiment. Rubric items B7 and A11.

From Fig. 2, we see that the apparatus group did significantly better at finding a pattern. Their scores are distributed evenly across the scale while in the video group, 70% of the students scored inadequately for finding a pattern in the data. The video group was significantly better at representing data graphically. Bar graphs of the score distributions of the two groups for rubric items B7 and A11 are shown in Fig. 3.

2. Testing experiments

The scores of the reports of the testing experiments show that the groups were significantly different in identifying assumptions (C5) and when making a judgment about the hypothesis under test (C8). We present the rubric items for those abilities (Table VII) and bar graphs showing the distributions of students' responses (Fig. 4).

Figure 4 shows why the two groups are different in identifying assumptions. It looks like the majority of the apparatus group did not do it at all. Concerning making a judgment, we see that half of the students in both groups did not write anything about their judgment. While there were many more reports in the apparatus group that received the score "needs improvement," and almost none in the "inadequate" category, only video students received "adequate" scores.

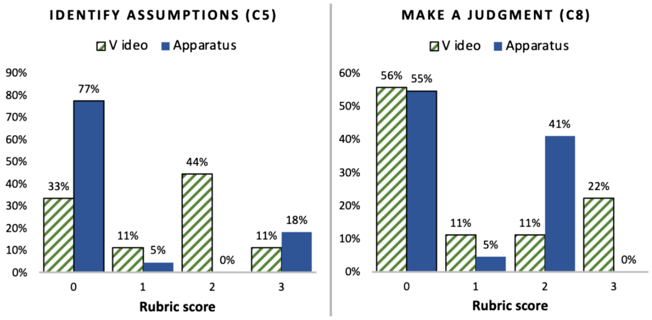


FIG. 4. Comparing abilities involved in designing and conducting a testing experiment. Rubric items C5 and C8.

3. Application experiments

The same situation repeats when the students conduct application experiments. On one of the rubric items, the apparatus group is significantly better on another, the video group. The rubrics for which we found significant differences assess students' ability to evaluate the results using a second method and to devise a mathematical procedure to solve the problem (Table VIII). Scoring the reports for the ability to identify assumptions in the mathematical procedure shows a significant difference but the picture is unclear, we need to examine the bar graphs (Fig. 5) to understand what is going on.

Figure 5 shows that students in the video group were more successful at designing an independent experiment to determine the same quantity. However, we need to remind the reader that in the video condition, the second experiment was designed and videoed for the students, while for the apparatus condition, the students needed to design and conduct the experiment themselves. It is possible that they could not come up with a second design, skipped over it, or ran out of time describing it in their reports. The data for the mathematical procedure slightly favor the apparatus group. We see from the graphs that the apparatus group is worse at identifying assumptions in the mathematical procedure. Most of the students did not identify assumptions at all. In contrast, the majority of the video group attempted to identify assumptions, but those assumptions were irrelevant or incorrect. This difference might be a result of the fact that the video group had text boxes to fill out while the apparatus group had to address assumptions in their report writing without the same level of prompting.

4. Experimental uncertainties

It is interesting to see that one ability—the ability to collect and analyze data (rubric G) clearly favored the apparatus group. They were able to identify the sources of uncertainty better, and while they were not strong in evaluating uncertainties, they are significantly better than the video group. They were also better in minimizing uncertainties and in data analysis. The rubrics used for scoring are shown below (Table IX). The distributions of scores are shown in Fig. 6.

B. Exams

The results of the rubric scores of student responses to all the exam questions used in the study are shown in Tables X and XI. Table X shows the average score of each lab section in both Fall 2018 and Spring 2019 for each question. Each question was scored using a rubric like the ones shown in Tables IV and V. Two coders, who were blind to what lab section the students were in, independently scored a small sample of each exam question. When interrater reliability of greater than 80% was achieved, one of the coders scored the remainder of the questions.

Table XI shows the results of a 1 × 4 ANOVA in which we treated each lab section as a separate experimental condition and each semester as a separate

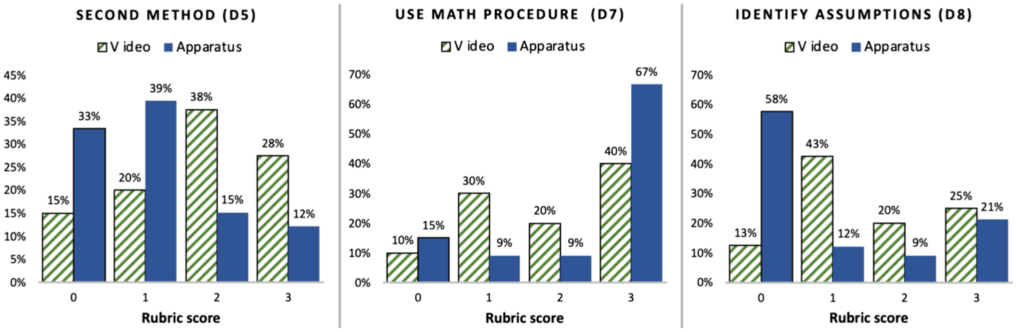


FIG. 5. Comparing abilities involved in designing and conducting an application experiment. Rubric items D5, D7, and D8.

TABLE VII. Rubric items used for the analysis of testing experiment reports.

Ability	Missing (0)	Inadequate (1)	Needs improvement (2)	Adequate (3)
Is able to identify the assumptions made in making the prediction (C5)	No attempt is made to identify any assumptions.	An attempt is made to identify assumptions, but the assumptions are irrelevant or are confused with the hypothesis.	Relevant assumptions are identified but are not significant for making the prediction.	Sufficient assumptions are correctly identified and are significant for the prediction that is made.
Is able to make a reasonable judgment about the hypothesis (C8)	No judgment is made about the hypothesis.	A judgment is made but is not consistent with the outcome of the experiment.	A judgment is made, is consistent with the outcome of the experiment, but assumptions are not taken into account.	A judgment is made, consistent with the experimental outcome, and assumptions are taken into account.

experiment. Note that we did not make any comparisons between semesters since students were not randomly assigned to a semester and so it is not valid to assume the two semesters were drawn from the same population.

The intent of the energy question was to test whether there was any inherent difference between the lab sections by asking a question about energy bar charts, a topic unrelated in terms of both content and scientific abilities to anything in the experimental intervention. In neither semester were there any significant differences between conditions. The Newton's laws question assessed conceptual understanding of Newton's second and third laws, which was related to one of the labs that the students did with apparatus and video. Again, there was no significant difference between conditions in either semester. Finally, we compared student scores on the question that assessed students' abilities to analyze and interpret data. Since the content area of the question (the dependence of the period of a pendulum swing on the amplitude of that swing) was not part of the intervention, this question would indicate if any of the treatments affected students' ability to analyze data, estimate uncertainties, and draw conclusions from data. Here there was a significant difference between conditions in the Fall 2018 experiment, but not in the Spring 2019 experiment.

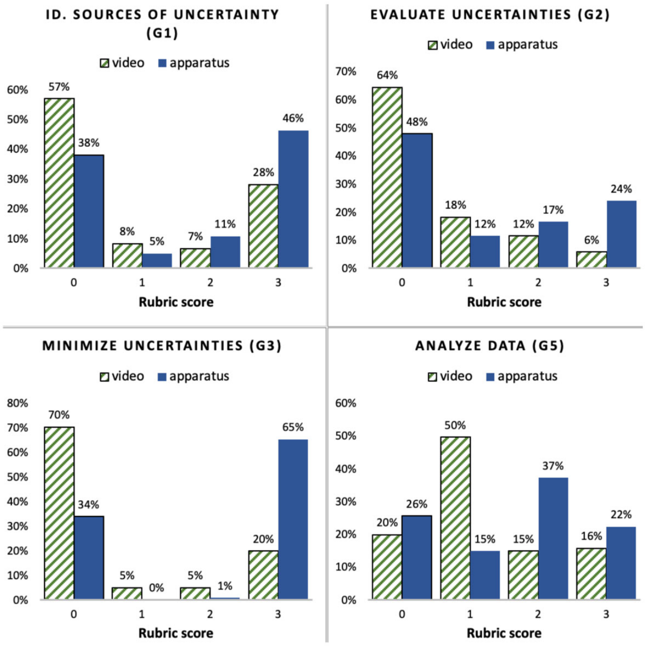


FIG. 6. Comparing abilities to collect and analyze experimental data during video-based and apparatus-based labs. Rubric items G1, G2, G3, and G5.

VI. DISCUSSION AND IMPLICATIONS FOR INSTRUCTION

We were able to design lab activities for the students using video-based experiments that in all cases but one replicated activities with physical apparatus. An important issue here is that the activities were designed for the ISLE approach where one of the main aspects of student experimental work is experimental design. We found that it was possible to design video-based activities that mimicked the process of design of experiments by the students. This opportunity was provided by the unique features of Pivot Interactives where students could select from a matrix of videos (thus learning to control variables in experimental design) and use overlaid measurement tools to make measurements (mimicking the same sort of measurement decisions students have to make when working with physical apparatus). The only experiment that was different for the students in two conditions was the observational experiment where students developed a mathematical expression for the radial acceleration of an object moving in a circle at a constant speed. Here the video condition actually was preferential as the students could vary each variable separately, whereas, using physical equipment, this experiment is difficult for students as it either requires a specific expensive apparatus developed exactly for this

TABLE VIII. Rubric items used for the analysis of application experiment reports.

Ability	Missing (0)	Inadequate (1)	Needs improvement (2)	Adequate (3)
Is able to evaluate the results by means of an independent method (D5)	No attempt is made to evaluate the consistency of the result using an independent method.	A second independent method is used to evaluate the results. However, there is little or no discussion about the differences in the results due to the two methods.	A second independent method is used to evaluate the results. The results of the two methods are compared correctly using experimental uncertainties. But there is little or no discussion of the possible reasons for the differences when the results are different.	A second independent method is used to evaluate the results and the evaluation is correctly done with the experimental uncertainties. The discrepancy between the results of the two methods and possible reasons are discussed.
Is able to choose a productive mathematical procedure for solving the experimental problem (D7)	Mathematical procedure is either missing or the equations written down are irrelevant to the design.	A mathematical procedure is described, but is incorrect or incomplete, due to which the final answer cannot be calculated. Or units are inconsistent.	Correct and complete mathematical procedure is described but an error is made in the calculations. All units are consistent.	Mathematical procedure is fully consistent with the design. All quantities are calculated correctly with proper units. Final answer is meaningful.
Is able to identify the assumptions made in using the mathematical procedure (D8)	No attempt is made to identify any assumptions.	An attempt is made to identify assumptions, but the assumptions are irrelevant or incorrect for the situation.	Relevant assumptions are identified but are not significant for solving the problem.	All relevant assumptions are correctly identified.

purpose or using clunky equipment, collecting large amounts of messy data. In our case, the students in the apparatus condition came up with the same relationship through the analysis of a thought experiment [41].

More specifically, our research goal was to answer the following two questions:

1. What are the differences in how students develop scientific abilities when experimenting with videos versus apparatus in ISLE-based labs?
2. How does experimenting with video affect student learning of normative physics concepts in an ISLE-based course such as (for example) Newton's second and third laws as compared to students who learn the same physics ideas by experimenting with apparatus?

We will answer each question first and then provide a summary.

A. What are the differences in how students develop scientific abilities when experimenting with videos versus apparatus in ISLE-based labs?

From the analysis of students' lab reports using scientific abilities rubrics, we can conclude that there are scientific abilities whose development is better supported by using physical apparatus and there are some abilities that are better supported when students do video experiments. While we found only some random differences in students' reports of different types of experiments (on some rubrics, the apparatus group did better and on some, the video group), these differences were probably a consequence of the format of the lab itself. However, there is one rubric (G-Data collection and analysis) on which students who did apparatus labs showed better results on almost every rubric item. This finding is robust as we scored every lab report we analyzed using rubric G, therefore, the data sample is much larger for this rubric. We found that, compared to the video condition, the apparatus condition students could better identify, analyze, and describe how to minimize experimental uncertainties and analyze data in the lab reports we scored. This is perhaps not that surprising. *A priori*, we suspected that the "messier" data provided by apparatus-based experimentation would offer more opportunities for students to think about sources of experimental uncertainty and how to minimize them. In contrast, when working with video-based experiments, it is harder for students to quantify random uncertainties.

TABLE IX. Rubric items used for the analysis of students' abilities to analyze data and deal with experimental uncertainties.

Ability	Missing (0)	Inadequate (1)	Needs improvement (2)	Adequate (3)
Is able to identify sources of experimental uncertainty (G1)	No attempt is made to identify experimental uncertainties.	An attempt is made to identify experimental uncertainties, but most are missing, described vaguely, or incorrect.	Most experimental uncertainties are correctly identified.	All experimental uncertainties are correctly identified.
Is able to evaluate specifically how identified experimental uncertainties may affect the data (G2)	No attempt is made to evaluate experimental uncertainties.	An attempt is made to evaluate experimental uncertainties, but most are missing, described vaguely, or incorrect. Or only absolute uncertainties are mentioned. Or the final result does not take the uncertainty into the account.	The final result does take the identified uncertainties into account but is not correctly evaluated. The weakest link rule is not used or is used incorrectly.	The experimental uncertainty of the final result is correctly evaluated. The weakest link rule is used appropriately and the choice of the biggest source of uncertainty is justified.
Is able to describe how to minimize experimental uncertainty and actually do it (G3)	No attempt is made to describe how to minimize experimental uncertainty and no attempt to minimize is present.	A description of how to minimize experimental uncertainty is present, but there is no attempt to actually minimize it.	An attempt is made to minimize the uncertainty in the final result, but the method is not the most effective.	The uncertainty is minimized in an effective way.
Is able to analyze data appropriately (G5)	No attempt is made to analyze the data.	An attempt is made to analyze the data, but it is either seriously flawed or inappropriate.	The analysis is appropriate but it contains minor errors or omissions.	The analysis is appropriate, complete, and correct.

On the exam question where students were asked to analyze a dataset (see Table XI), there was one significant difference between two of the conditions: lab section 3 (treatment VVAAA) and lab section 4 (treatment VVVVV) in the fall of 2018. Section 4 performed significantly better. However, the difference could have been due to an idiosyncratic difference between the two groups or random chance, as it did not appear in the spring of 2019. We conclude that, on the whole, students in all conditions did equally well at data analysis by the end of the semester. This begs the question: Why do we see such a marked difference in the labs but not in the final exam? There are two possibilities. (1) Prior research [14] has shown that students do learn how to analyze data and calculate uncertainties from video-based experiments. (2) Students in the video conditions also did many experiments with apparatus as well, so they likely became more adept at analyzing and interpreting noisy experimental data from those lab experiences.

TABLE X. Average scores for all exam questions analyzed in the study, broken down by section and semester.

Section		1	2	3	4
Treatment		5A	2A3V	2V3A	5V
Fall '18	Energy Q.	62%	67%	64%	56%
	N2/N3 Q.	33%	36%	39%	25%
	Data Q.	31%	47%	30%	53%
Spring '19	Energy Q.	57%	61%	45%	60%
	N2/N3 Q.	48%	51%	54%	64%
	Data Q.	36%	44%	41%	42%

TABLE XI. All results of a one-way ANOVA with four independent conditions applied to each exam question, treating each semester as an independent experiment.

	Energy Q.		N2/N3 Q.		Data Q.	
	F18	S19	F18	S19	F18	S19
F[a]	0.89	1.45	1.03	1.09	3.7	0.35
p	0.45	0.23	0.38	0.36	0.016[b]	0.79
N	84	90	86	93	72	62

[a] Three degrees of freedom throughout (df = 3).

[b] Tukey's test reveals that the only statistically significant difference is between lab section 3 (treatment VVAAA) and lab section 4 (treatment VVVVV). Section 3 scored 30% and section 4 scored 53%, $p < 0.05$. All other pairwise comparisons were not significant.

At the same time, there is another ability (rubric item A11, ability to plot a graph), where students in the video condition showed better results than apparatus students (see Fig. 3). While this rubric was used for only one lab, Newton's second law, this finding deserves special attention. This is the lab where the students need to make acceleration vs mass and acceleration vs force graphs and from those come up with the direct and inverse relationships between acceleration, the force exerted on an object and its mass. The fact that almost 80% of the students in the video condition were able to do it perfectly shows how beneficial precise data are for developing mathematical relations from that data.

The above discussion shows that it is beneficial for the students to have both types of labs when both follow the ISLE philosophy. The labs using apparatus help students identify and analyze experimental uncertainties, video-based labs help students develop relationships based on data when data collection is messy and tedious. Asking students to experimentally develop mathematical relations like the dependence of the centripetal acceleration on speed and radius, with physical apparatus, can require a large time investment for students and a large financial investment from the school ($325 per setup as of the time of writing this paper [42]). Video experiments definitely save time and funds when it comes to complicated data collection when the students need to develop mathematical relationships in which some aspects are nonlinear.

B. How does experimenting with video affect student learning of normative physics concepts in an ISLE-based course?

The data presented in Sec. V show that the students in different conditions were indistinguishable on the exam questions unrelated to the content of the labs where the conditions were different (apparatus and video) and also on the questions related to the content of the labs (Newton's second and third laws) where the conditions were different. While performance on a single exam question does not suffice to draw any general conclusions, our results lend support to the claim that substituting some of the apparatus labs with the video analysis labs does not seem to affect the quality of student learning of normative content when students are learning physics through the ISLE approach.

C. Summary

It does appear from our data that video-based experiments offer certain affordances and constraints in terms of supporting students' development of scientific abilities. Instructors engaged in designing labs for students need to make conscious choices when deciding whether a specific lab can be or even should be substituted with high-quality videos of a similar experiment. Some of us are strong proponents of using regular equipment in student labs and some lack funds to purchase necessary equipment. Our study suggests that a balanced mix of both kinds of experiments might benefit the students the most in terms of fostering the growth of all scientific abilities.

Additionally, as online and asynchronous learning continues to expand, there is potential in combining video-based experiments using expensive apparatus with at-home experiments using everyday objects as a means to overcome the identified weakness regarding identification, analysis, and minimization of experimental uncertainties. The combination of the two might even reduce the need for traditional apparatus-based laboratory facilities while providing online students with more meaningful learning experiences.

D. Limitations and future work

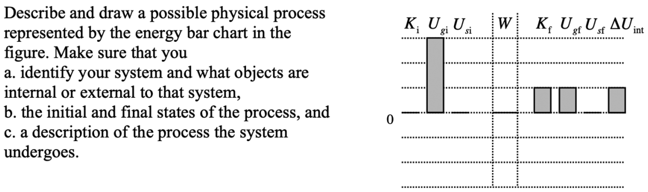
Our study has several limitations. First, we only substituted 5 labs out of 15 with video experiments, therefore, it is difficult to speculate what would happen if the students did all 15 labs using video-based experiments instead of using apparatus. Second, the project was conducted in a state devastated by fires at the time of the study. This could have affected student growth and development as some of them did not have a place to live and work. Third, our relatively small exam dataset means that we be cautious about the generalizability of our conclusions from that dataset.

Future research could test some of our hypotheses by having students conduct an entire semester of experimentation with video rather than physical apparatus. The robustness of our findings could be tested with a larger sampling of students across multiple institutions.

ACKNOWLEDGMENTS

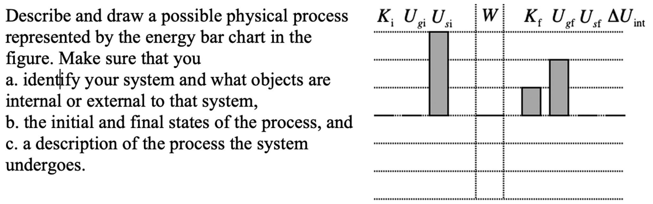
This work is supported by NSF Grant No. DUE-1726249. Viewpoints expressed here are those of the authors and do not reflect the views of NSF. We thank Dedra Demaree for her continuous feedback on the design of Pivot lab activities and the overall support of the project. Thank you Aaron Warren and Michael Kagan for supporting our project.

APPENDIX: EXAM QUESTIONS



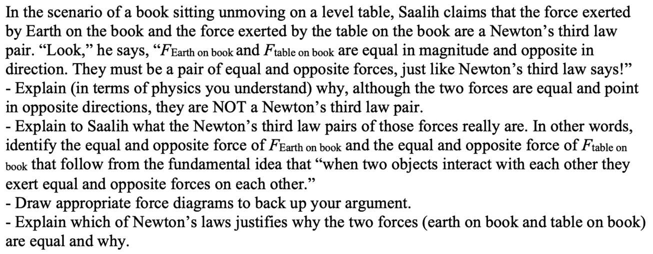
Describe and draw a possible physical process represented by the energy bar chart in the figure. Make sure that you
a. identify your system and what objects are internal or external to that system,
b. the initial and final states of the process, and
c. a description of the process the system undergoes.

FIG. 7. Fall 2018 energy bar chart question assessing students' representation abilities in the context of energy.



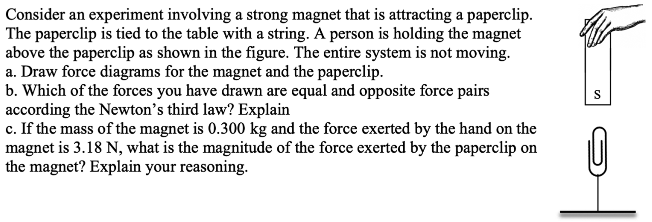
Describe and draw a possible physical process represented by the energy bar chart in the figure. Make sure that you
a. identify your system and what objects are internal or external to that system,
b. the initial and final states of the process, and
c. a description of the process the system undergoes.

FIG. 8. Spring 2019 energy bar chart question assessing students' representation abilities in the context of energy.



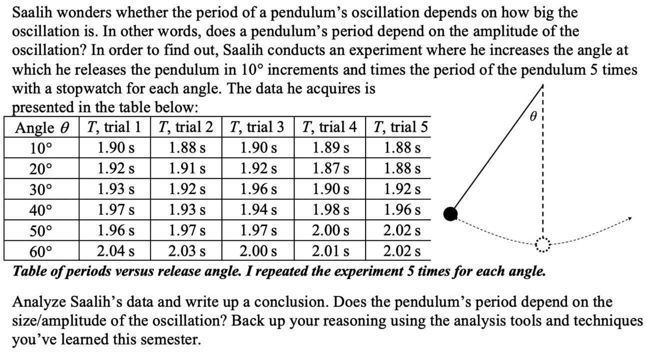
In the scenario of a book sitting unmoving on a level table, Saalih claims that the force exerted by Earth on the book and the force exerted by the table on the book are a Newton's third law pair. "Look," he says, "$F_{\text{Earth on book}}$ and $F_{\text{table on book}}$ are equal in magnitude and opposite in direction. They must be a pair of equal and opposite forces, just like Newton's third law says!"
- Explain (in terms of physics you understand) why, although the two forces are equal and point in opposite directions, they are NOT a Newton's third law pair.
- Explain to Saalih what the Newton's third law pairs of those forces really are. In other words, identify the equal and opposite force of $F_{\text{Earth on book}}$ and the equal and opposite force of $F_{\text{table on book}}$ that follow from the fundamental idea that "when two objects interact with each other they exert equal and opposite forces on each other."
- Draw appropriate force diagrams to back up your argument.
- Explain which of Newton's laws justifies why the two forces (earth on book and table on book) are equal and why.

FIG. 9. Newton's second or third law question used (Fall 2018).



Consider an experiment involving a strong magnet that is attracting a paperclip. The paperclip is tied to the table with a string. A person is holding the magnet above the paperclip as shown in the figure. The entire system is not moving.
a. Draw force diagrams for the magnet and the paperclip.
b. Which of the forces you have drawn are equal and opposite force pairs according the Newton's third law? Explain
c. If the mass of the magnet is 0.300 kg and the force exerted by the hand on the magnet is 3.18 N, what is the magnitude of the force exerted by the paperclip on the magnet? Explain your reasoning.

FIG. 10. Newton's second or third law question used (Spring 2019).



Saalih wonders whether the period of a pendulum's oscillation depends on how big the oscillation is. In other words, does a pendulum's period depend on the amplitude of the oscillation? In order to find out, Saalih conducts an experiment where he increases the angle at which he releases the pendulum in 10° increments and times the period of the pendulum 5 times with a stopwatch for each angle. The data he acquires is presented in the table below:

Angle θ	T, trial 1	T, trial 2	T, trial 3	T, trial 4	T, trial 5
10°	1.90 s	1.88 s	1.90 s	1.89 s	1.88 s
20°	1.92 s	1.91 s	1.92 s	1.87 s	1.88 s
30°	1.93 s	1.92 s	1.96 s	1.90 s	1.92 s
40°	1.97 s	1.93 s	1.94 s	1.98 s	1.96 s
50°	1.96 s	1.97 s	1.97 s	2.00 s	2.02 s
60°	2.04 s	2.03 s	2.00 s	2.01 s	2.02 s

Table of periods versus release angle. I repeated the experiment 5 times for each angle.

Analyze Saalih's data and write up a conclusion. Does the pendulum's period depend on the size/amplitude of the oscillation? Back up your reasoning using the analysis tools and techniques you've learned this semester.

FIG. 11. Data analysis question used to assess students' abilities to analyze data and make a data-based judgment.

[1] E. Etkina and A. Van Heuvelen, Investigative science learning environment—a science process approach to learning physics, in *Research-Based Reform of University Physics*, edited by E. F. Redish and P. J. Cooney (2007), Vol. 1.

[2] R. K. Thornton and D. R. Sokoloff, Learning motion concepts using real-time microcomputer-based laboratory tools, Am. J. Phys. **58**, 858 (1990).

[3] N. D. Finkelstein, W. K. Adams, C. J. Keller, P. B. Kohl, K. K. Perkins, N. S. Podolefsky, S. Reid, and R. LeMaster, When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment, Phys. Rev. ST Phys. Educ. Res. **1**, 010103 (2005).

[4] R. E. Mayer, *Multimedia Learning* (Cambridge University Press, Cambridge, London, 2009), 2nd ed.

[5] T. Stelzer, G. Gladding, J. P. Mestre, and D. T. Brookes, Comparing the efficacy of multimedia modules with traditional textbooks for learning introductory physics content, Am. J. Phys. **77**, 184 (2009).

[6] N. G. 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).

[7] 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).

[8] E. Etkina, A. Van Heuvelen, S. White-Brahmia, D. T. Brookes, M. Gentile, S. Murthy, D. Rosengrant, and A. Warren, Scientific abilities and their assessment, Phys. Rev. ST Phys. Educ. Res. **2**, 020103 (2006).

[9] E. Etkina, A. Karelina, and M. Ruibal-Villasenor, How long does it take? A study of student acquisition of scientific abilities, Phys. Rev. ST Phys. Educ. Res. **4**, 020108 (2008).

[10] D. T. Brookes and E. Etkina, Physical phenomena in real time, Science **330**, 605 (2010).

[11] The developers of Pivot Interactives subsequently introduced the coLAB synchronous collaboration feature in Pivot partially based on feedback from the research stemming from this project.

[12] W.-M. Roth and D. Lawless, Scientific investigations, metaphorical gestures, and the emergence of abstract scientific concepts, Learn. Instr. **12**, 285 (2002).

[13] M. Kardas and E. O'Brien, Easier seen than done: Merely watching others perform can foster an illusion of skill acquisition, Psychol. Sci. **29**, 521 (2018).

[14] M. Vonk, P. Bohacek, C. Militello, and E. Iverson, Developing model-making and model-breaking skills using direct measurement video-based activities, Phys. Rev. Phys. Educ. Res. **13**, 020106 (2017).

[15] E. Brewe, E. Etkina, and K. Neumann, Holistic learning environments, in *The International Handbook of Physics Education Research: Teaching Physics* (AIP Publishing LLC, Melville, NY, 2023), pp. 1–8.

[16] E. Etkina, A. Van Heuvelen, D. T. Brookes, and D. Mills, Role of experiments in physics instruction—a process approach, Phys. Teach. **40**, 351 (2002).

[17] E. Etkina, Millikan award lecture: Students of physics—listeners, observers, or collaborative participants in physics scientific practices? Am. J. Phys. **83**, 669 (2015).

[18] E. Etkina, D. T. Brookes, and G. Planinsic, *Investigative Science Learning Environment When Learning Physics Mirrors Doing Physics* (Morgan & Claypool, San Rafael, CA, 2019).

[19] D. T. Brookes, E. Etkina, and G. Planinsic, Implementing an epistemologically authentic approach to student-centered inquiry learning, Phys. Rev. Phys. Educ. Res. **16**, 020148 (2020).

[20] N. G. Holmes, Structured quantitative inquiry labs: Developing critical thinking in the introductory physics laboratory, Ph.D. thesis, University of British Colombia, Vancouver, BC, 2014.

[21] R. Lippmann Kung, Teaching the concepts of measurement: An example of a concept-based laboratory course, Am. J. Phys. **73**, 771 (2005).

[22] E. Etkina, S. Murthy, and X. Zou, Using introductory labs to engage students in experimental design, Am. J. Phys. **74**, 979 (2006).

[23] A. Karelina and E. Etkina, When and how do students engage in sense-making in a physics lab? AIP Conf. Proc. **883**, 93 (2007).

[24] E. Etkina, A. Karelina, S. Murthy, and M. Ruibal-Villasenor, Using action research to improve learning and formative assessment to conduct research, Phys. Rev. ST Phys. Educ. Res. **5**, 010109 (2009).

[25] D. Bugge and E. Etkina, Reading between the lines: Lab reports help high school students develop scientific abilities, presented at PER Conf. 2016, Sacramento, CA, http://doi.org/10.1119/perc.2016.pr.008.

[26] J. Rutberg, M. Malysheva, and E. Etkina, Impact of ISLE-based labs in courses with traditional lecture, presented at PER Conf. 2019, Provo, UT, http://doi.org/10.1119/perc.2019.pr.Rutberg.

[27] A. Karelina, E. Etkina, P. Bohacek, M. Vonk, M. Kagan, A. R. Warren, and D. T. Brookes, Comparing students' flow states during apparatus-based versus video-based lab activities, Eur. J. Phys. **43**, 045701 (2022).

[28] See https://www.pivotinteractives.com/.

[29] See https://app.pivotinteractives.com/activities/5a1f29502-ace2c002065f83b/preview.

[30] See https://app.pivotinteractives.com/activities/5a74d-8287cff71001a355c46/preview.

[31] See https://app.pivotinteractives.com/activities/5a7df0-cbfd95f4001a5319b3/preview.

[32] See https://app.pivotinteractives.com/activities/5a6deec-203db79001a222161/preview.

[33] See https://app.pivotinteractives.com/activities/5a81ac5b-11d1450020d278e9/preview.

[34] See https://app.pivotinteractives.com/activities/5a32f48f-2201480020a3612c/preview.

[35] See https://app.pivotinteractives.com/activities/5a3d85-daadea50002024cf95/preview.

[36] See https://app.pivotinteractives.com/activities/5b118998-db66f60020a64b2a/preview.

[37] See https://app.pivotinteractives.com/activities/5a87c0e0c-8e2a3001a168f0b/preview.

[38] Unavailable without a pivot account.

[39] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.19.020158 for a side-by-side comparison of two equivalent experiments, one apparatus-based, one video-based, related to the period of oscillation of an object on a spring.

[40] See https://sites.google.com/site/scientificabilities/rubrics.

[41] E. Etkina, G. Planinsic, and A. Van Heuvelen, *College Physics: Explore and Apply* (Pearson, San Francisco, CA, 2019), 2nd ed., pp. 123–125.

[42] https://www.vernier.com/product/go-direct-centripetal-force-apparatus/.