

Student engagement with modeling in multiweek student-designed lab projects

Victoria Borish^{1,2,*}, Jessica R. Hoehn^{1,2} and H. J. Lewandowski^{1,2}

Department of Physics, University of Colorado, Boulder, Colorado 80309, USA
JILA, National Institute of Standards and Technology and University of Colorado,
Boulder, Colorado 80309, USA

 (Received 23 December 2021; accepted 21 October 2022; published 21 November 2022)

Modeling, which is the process of constructing, testing, and refining models, is an important skill in experimental physics, and thus a learning goal of many physics laboratory classes. One promising approach to help students develop modeling skills is to incorporate multiweek student-designed projects into lab courses. In order to assess the potential benefits of these projects in enhancing students' modeling abilities, we analyzed projects from three upper-division lab courses at different institutions. By looking at written student coursework and student interviews, we investigated which parts of the modeling process the students from each project undertook, and how this engagement in modeling differed depending on features of the projects. The projects in our dataset varied widely, showing evidence of different ways students engaged with model construction and revisions. We observed that the features of the projects, such as the goal of the project and the complexity of the required apparatus, were associated with the ways in which the students constructed models and enacted revisions. This has implications for how instructors may choose to frame and structure courses with student-designed lab projects.

DOI: [10.1103/PhysRevPhysEducRes.18.020135](https://doi.org/10.1103/PhysRevPhysEducRes.18.020135)

I. INTRODUCTION

The process of modeling is central to all sciences, and is therefore a main learning goal of many physics laboratory classes [1]. Models are abstract representations of real-world systems and are used to explain and predict phenomena. Modeling, the iterative process of constructing, testing, and refining models [2–4], is part of many tasks performed by experimental physicists [5]. Developing modeling skills has been identified as an important part of physics instruction for decades [6], and there has been a recent emphasis towards including modeling as a learning goal for undergraduate physics programs overall [7], and lab courses in particular [1].

One way physics instructors hope to teach their students modeling skills is through multiweek student-designed projects in lab courses. In these projects, students have the opportunity to perform an entire experiment, or a significant portion of it, from the conception of a research question to the presentation of results. The structure and implementation of these projects varies widely, with the possibility for students to formulate a scientific question, build an apparatus, take data, compare

data with a theoretical model, consider plausible causes if a discrepancy occurs, and iterate on these different steps until a satisfactory result can be presented. Engaging with the iterative nature of the modeling process is often a key component of these projects [8,9].

There are many documented benefits of student-designed projects, yet the extent to which they can be used to effectively teach modeling skills has not been investigated. Student-designed projects have been shown to improve student learning [10] and provide students authentic experimentation experiences [11–13]. These projects also allow students to develop a sense of ownership over their work, which has the potential to improve students' motivation and lead to persistence to remain in the field [14–17]. Additionally, they have improved student views about experimental physics [18], including student enthusiasm towards, and confidence in, performing experiments [19]. Because of the success of student-designed projects for those outcomes, it is worth investigating other potential benefits, such as students' improved modeling skills and understanding of the iterative nature of experimentation.

Our goal is to demonstrate various possible ways students may engage in the iterative modeling process while working on multiweek student-designed projects. We report findings of a qualitative study analyzing students' project artifacts (lab notebooks, proposals, reports, and presentations), written student reflections, and student interviews from three advanced lab courses at different institutions. We use the Experimental Modeling Framework (EMF) [4,20] as a tool in our thematic coding

*victoria.borish@colorado.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

analysis to investigate which elements of modeling the students demonstrate in their coursework. We consider not only whether these elements are demonstrated in each project, but also if there are associations between different elements of the EMF and features of the projects. Understanding how students may engage in different parts of the modeling process and how this depends on the features of the project can help instructors decide how best to frame these projects in their lab courses.

Our first research question directly investigates how students engage in modeling in each project:

RQ1: How do students engage in the process of modeling during multi-week student-designed projects?

To answer this question, we examine which elements of the EMF the students demonstrate in each project. We focus on the following subquestions:

- To what extent do students engage in modeling as defined by the EMF?
- Which elements of the EMF are most and least commonly demonstrated across all of the projects?
- What variations exist in the ways students working on different projects engage with modeling, as defined by the EMF?

After investigating the variation of student engagement in modeling by project, we examine the interplay between this engagement in modeling and features of the projects. Our second research question is

RQ2: Which features of multiweek student-designed projects are associated with the varied ways the students engage in modeling?

To explore this question, we look at overall trends across project features as well as specific instances of modeling demonstrated by students in individual projects. We focus on the following two subquestions about model construction and iteration, both of which are common learning goals and elements of the EMF with a large variation in student engagement:

- Which project features are associated with constructing models?
- Which project features are associated with different kinds of revisions?

The rest of this paper is organized as follows. Section II provides the necessary background describing prior research about both multiweek student-designed projects and modeling, with an emphasis on the EMF, which we use as a basis for our coding analysis. We then describe the methods of our study in Sec. III, including the courses, data sources, coding analysis, and limitations. Section IV discusses details of the specific projects we analyze, as well as the project features and modeling codes we assign them. Some of these codes are emergent from the coding process and therefore are a result in and of themselves, indicating how the students engaged in modeling and ways in

which the projects differed from one another. In Sec. V, we present our results about general trends across all projects, addressing both RQ1 and RQ2. We discuss our findings in Sec. VI, using student quotes to elucidate our claims and understand the role model construction played within different projects. We present implications for instruction based on these results in Sec. VII and conclude with an outlook for future research in Sec. VIII.

II. BACKGROUND

The two primary areas of research relevant to this study are multiweek student-designed projects, which we generalize to include all lab coursework involving student decision making, and modeling of experiments. In order to situate our work in a larger research context, we first review the previously demonstrated benefits of providing students opportunities to make decisions in labs in Sec. II A, with a focus on the documented learning and affective gains from multiweek student-designed projects. In Sec. II B, we briefly discuss prior research on modeling in the context of physics education. We then present the EMF, the theoretical framework we use in our study to evaluate how the students engage in the modeling process.

A. Student decision making in labs

There is a trend in the physics education community to shift away from traditional prescriptive labs towards labs that allow students to make decisions about various parts of the experimental process [21–24]. There is a large range of possible decisions students can make, from deciding on their research question to designing the experimental procedure, to choosing how to analyze the data. Although in traditional labs students follow a detailed set of instructions, it is possible to allow for student decision making over any component of the lab or even multiple components at once, as is the case for multiweek student-designed projects [25]. These opportunities to practice decision making help students learn the problem-solving skills used by expert physicists [26,27].

Prior research has shown a variety of benefits to students arising from making decisions in labs, including improving students' views of experimental physics and providing opportunities for authentic engagement in scientific practices. Wilcox and Lewandowski showed that open-ended activities, whether they were multiweek projects or shorter activities, led to students having more expertlike views about experimental physics [18]. The reasoning behind some students' epistemological views were investigated in Ref. [28], where it was shown that some of the nonexpertlike ways students viewed physics experiments came from experiences with labs in which they confirmed previously known results. Irving and Sayre showed that communities of practice that are similar to those of practicing physicists formed in classes where the students performed multiweek

experiments and were responsible for doing much of the troubleshooting themselves [29]. Other work showed that affording more opportunity for students' agency in labs allows them to practice authentic scientific decision making [30], develop inquiry skills [31], and develop core experimental skills [32]. Additionally, the ability to think critically and make decisions can lead to a higher rate of student enjoyment of the course [32–34].

Multiweek projects often allow students the opportunity to make many kinds of experimental decisions, and have been shown to improve student learning, increase engagement in authentic physics experiences, and improve student affect. Juma *et al.* showed that projects with minimal structure increase students' self-reported learning for both concepts and experimental skills [10]. Holmes and Wieman showed that students in design-based labs performed more of the cognitive tasks associated with experimental physicists than those in traditional labs [11], and Hoehn and Lewandowski showed that these projects provide students authentic writing experiences [12,13]. There have been results demonstrating affective benefits as well, such as increased feelings of ownership [14–17] and improvements in students' enjoyment of, and confidence in, performing physics experiments [19]. Here, we investigate another potential benefit of multiweek student-designed projects, namely, the way they provide students opportunities to engage in modeling.

B. Experimental Modeling Framework

There has been extensive research on modeling throughout physics and other education fields, yet there is still work to be done to understand how best to help students develop modeling skills in physics lab courses. Early work by Hestenes proposed focusing on modeling as a central part of physics instruction [6]. Since then, others have created other frameworks describing how physicists apply modeling [4,35] and designed curricula to teach modeling skills at the introductory level [36,37]. Recent work has begun to generate methods of evaluating students' use of modeling, such as by creating an analytic framework to characterize how students use a model to solve conceptual problems [38] and by designing large-scale assessments of certain aspects of model-based reasoning in introductory courses [39–41] and of the modeling process in upper-division optics and electronics labs [42–45]. Here, we probe student engagement in modeling not with a standardized closed-response assessment, but by examining advanced lab courses at three different institutions in depth to understand the ways in which students engage in modeling.

In order to investigate student engagement in modeling skills, we use the EMF. Zwickl *et al.* developed this framework in Ref. [4], and it is further expounded upon, along with a description of its initial applications, in Ref. [20]. It was first developed while transforming an advanced lab class, and it has been used to understand

student reasoning around modeling in think-aloud interviews in both optics and electronics contexts [4,46]. It has also been used as a tool to analyze student lab notebooks for evidence of model-based reasoning in an advanced electronics lab [47] and to investigate student views about the nature of models of experimental physics [48]. The framework has additionally inspired other course materials aimed at improving students' modeling skills [49,50]. Before the work presented here, the framework had not been applied as a tool to investigate student engagement in modeling for open-ended projects or experiments in fields outside of optics and electronics.

The EMF describes the iterative process of modeling applied to experimental physics and separates the process into tasks, as shown in Fig. 1. A key feature of this framework is the distinction between the physical system, which describes the phenomenon of interest, and the measurement system, which includes the components needed to collect and process data. Consideration of the model of the measurement system is crucial for most of experimental physics and thus makes this a suitable framework for use in upper-division physics lab courses.

The EMF consists of five main tasks: *Make measurements, construct models, make comparisons, propose causes, and enact revisions*. A measurement occurs when the physical system apparatus interacts with the measurement system apparatus. In order to compare the experimental data with a theoretical model, models of both the physical system and the measurement system need to be constructed. Once a comparison is made, if the experimental measurement agrees with the prediction to a level that is appropriate for the goals of the experiment, the process may stop. Otherwise, ideas about possible causes of this discrepancy can be generated, which will lead to revisions of the apparatus or the models. This process repeats until the agreement between the measurement and the prediction is deemed sufficient.

Key to the ability to make a comparison is the process of data analysis, where the raw data is converted into a form that can be compared with a prediction. Data analysis is illustrated in Fig. 1 by the arrows connecting the raw and interpreted data boxes. The exact method of data analysis will depend on both the model of the measurement system and the prediction. For our study, we consider data analysis its own task because it is distinct from, although also dependent on, the model of the measurement apparatus.

This framework may be used not only for the main experimental cycle where the comparison involves a measurement of the goal quantity, but also for troubleshooting any part of the system throughout the process [46]. We define the goal quantity as the measurement needed to answer the research question. For example, if the goal of an experiment is to measure the cross section of light absorption of a species of atom, then the cross section, or the raw data analyzed to obtain it, is considered the goal quantity. During

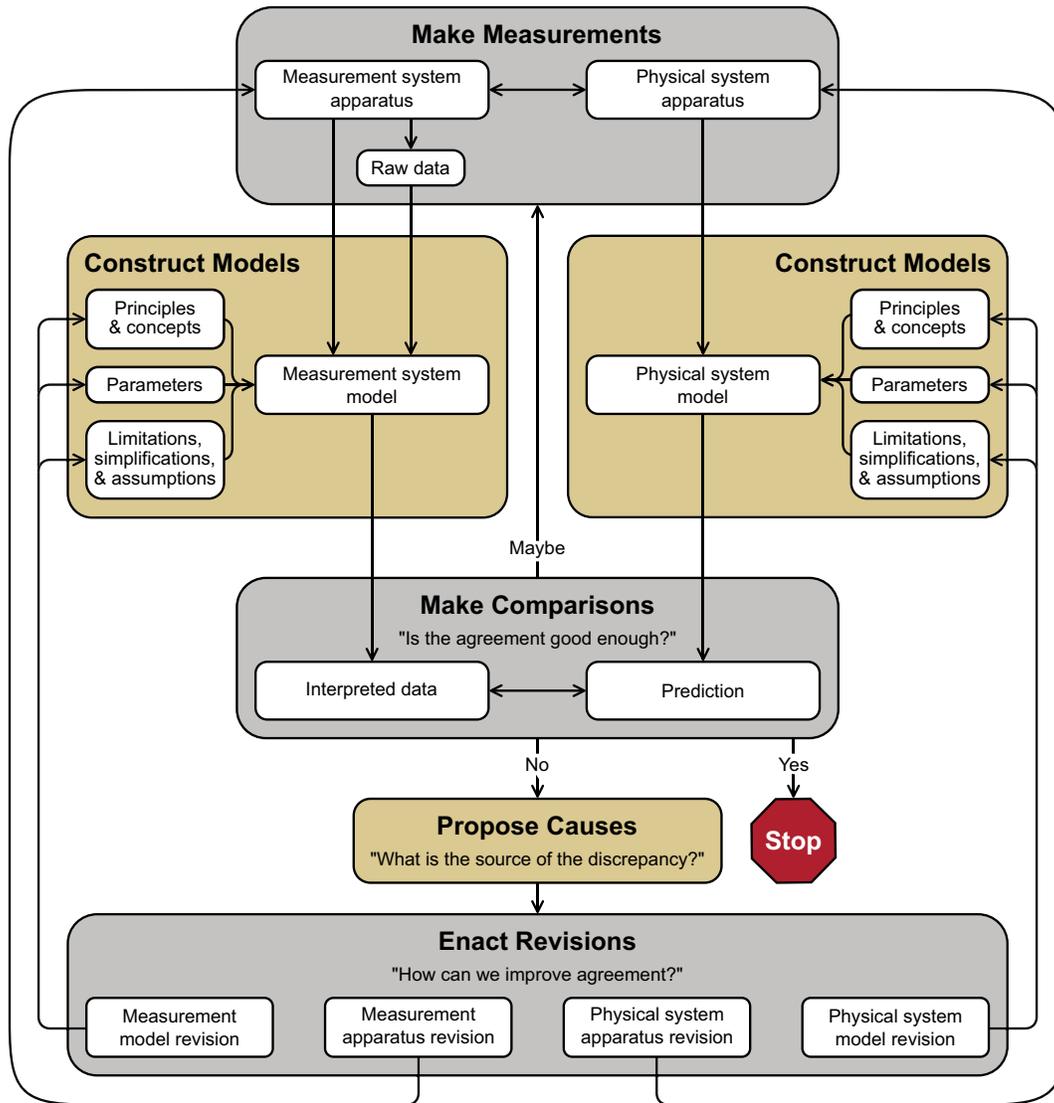


FIG. 1. Diagram of the Experimental Modeling Framework, which describes the process of modeling in experimental physics. This figure was first published in Ref. [42] and is a modified version of the original framework developed in Ref. [4]. The shaded boxes show the main tasks of modeling with the connecting arrows indicating the iterative nature of the modeling process.

the experimental process, there will likely be other measurements made and compared with other model predictions to ensure the apparatus is working as expected. These comparisons could be either quantitative or qualitative, and could lead to many different subcycles of model-based reasoning, all of which can be described by the EMF.

III. METHODS

In order to study student engagement in modeling in multiweek student-designed projects, we collected and coded project artifacts (proposals, reports, presentations, and lab notebooks), weekly written reflections, and interviews from students in advanced labs at three different institutions. We used the EMF as an analytic framework to code these data for the extent to which students engaged in

modeling and additionally coded features of the projects. This section describes the three courses from which we collected data, the data sources we analyzed, our coding analysis method, and limitations of the study, including those resulting from the COVID-19 pandemic. Additional details of our methodology can be found in Appendix A.

A. Course contexts

As part of a multiyear research project investigating effective practices for implementing multiweek student-designed projects in advanced lab classes, we partnered with instructors teaching advanced physics lab courses at three different institutions. These institutions span a variety of institution types and contexts, and all of the courses culminate in multiweek student-designed final projects.

TABLE I. Summary of the three courses from which we collected data.

	Course 1	Course 2	Course 3
Institution type	Small, private, predominantly white, liberal arts college	Large, public, master's-degree-granting, Hispanic-serving institution	Large, private, predominantly white, doctoral-degree-granting research university
Total number students	24 students	4 students	21 students
Typical level students	Juniors and seniors	Seniors	Juniors
Length of course	9 weeks	10 weeks	15 weeks
Course goals	Evaluate quality of data and compare with models, collaborate on and record making measurements and refining methodologies, communicate experimental results, and reflect on experimental physics	Work with advanced instrumentation, quantitative analysis, and realistic forms of scientific communication	Use contemporary experimental systems, design and carry-out projects, learn about proposal writing and evaluation, and present results
Effects of COVID	Entire course conducted remotely, final projects used equipment that could be shipped or found at students' homes	Almost entire course conducted before switch to remote learning	Switch to remote learning in middle of course, many students had to switch projects
Students per project	3 students	2 students	2–3 students
Length of final project	4 weeks	5 weeks	4 weeks
Final project preparation	Posted ideas to online forum with feedback from students and faculty, project proposals	Project proposals	White papers, half of those chosen and turned into project proposals with peer feedback

The data in this paper were taken from courses at these institutions in the winter or spring terms of 2020. Our in-depth understanding of the courses comes from instructor interviews, faculty online learning community meetings (in which all of the instructors participated), and course materials (e.g., syllabi). A summary of the courses is provided in Table I, and additional information can be found in Appendix A 1 and Ref. [12].

Course 1 was taught entirely remotely due to the COVID-19 pandemic, and all of the labs (two two-week-long structured labs and the final project), were adjusted so the students could conduct them at home. For the final projects, some of the students used materials found around their homes (e.g., various containers, yarn, cell phone sensors), while others were shipped materials that are commonly found in labs (e.g., laser pointers, force sensors, circuit elements). The students had time early on in the term to brainstorm ideas for their final projects, and then had four weeks to execute the projects, dividing up the work in various ways.

Course 2 also consisted of two two- to three-week-long structured labs followed by the final project. This course occurred almost entirely before the switch to remote learning caused by the COVID-19 pandemic, so all of the labs were conducted with students working in pairs at

their institution. The students received feedback on a project proposal early in the term and then had five weeks to complete the final project, which was cut short by only a couple of days due to the transition to remote learning.

Course 3 was longer than the other two courses and was interrupted in the middle by the COVID-19 pandemic. The first two-thirds of the course involved structured labs, as well as an extensive process of devising ideas and plans for the final projects. At the start of the four weeks dedicated to the projects, the course transitioned to remote learning. Several of the projects had to change, with the students accessing online experiments, bringing equipment home, going in to lab individually, or needing to switch to a literature review because they were not able to perform an experiment. The projects in this course were disrupted more than in the other two due to the timing of the pandemic.

B. Data sources

For each course, we analyzed several different data sources, as summarized in Table II. All of the courses required lab notebooks, final reports and/or presentations, and written student reflections, although the implementation of each varied by course. We additionally analyze the final project proposals for course 3 because it was the only course

TABLE II. Description of data sources used for each course. Dots indicate the source was not used for a specific course. Projects used shows the number of projects analyzed from that course compared with the total number of projects.

Data source	Course 1	Course 2	Course 3
Lab notebooks	Group electronic	Individual paper	Individual paper
Project proposal	Group
Final report	Group	Individual	...
Final presentation	Group
Reflections	Two sets paragraph-style	Five sets qualtrics	Six sets qualtrics
Interviews	3	2	5
Projects used	7/8	2/2	5/10

without written final reports and we did not have a full set of proposals from the other courses. To gain supplementary information about the students' projects, we interviewed 10 of the students across the three courses. Detailed descriptions of these data sources are provided in Appendix A 2.

After reviewing the data for all of the projects in the three courses, we decided to analyze the subset of projects which satisfied the following two criteria: (i) we had at least one student interview or complete lab notebook and (ii) the students performed, or began performing, an experiment. A complete lab notebook is defined as a notebook with entries through the final week of the time allocated for the project. We chose the first criterion because we felt we had insufficient information about the actions the students took while working on the projects without either a complete lab notebook or an interview. Additionally, we chose not to include the projects that had turned into literature reviews at the shift to remote learning because they are a distinct kind of lab activity, which is not the focus of our research questions. For the cases where the students changed their project partway through due to the COVID-19 pandemic, we analyzed only the data from after the change. Due to these criteria, we eliminated one project from course 1 and five projects from course 3. This left us 14 projects to analyze, each with a slightly different combination of the possible data sources.

C. Coding analysis

To investigate how students engaged in modeling, we coded all of the different data sources (lab notebooks, final reports and presentations, student reflections, and student interviews). All of the coded material was produced by the students. The codes and their definitions were refined through an iterative coding process. The final coding was done by one of the authors with frequent discussions about code definitions by all authors during the creation of the codebook.

Our codebook consists of two parts: one with codes describing the modeling tasks undertaken by the students

and the other with codes describing features of the projects. The modeling codes describe actions taken by the students, including both *a priori* codes coming directly from the main tasks of the EMF (Fig. 1) and emergent subcodes describing the specific ways the students engaged with some of these tasks. The feature codes are descriptions of the projects that are independent of (although possibly correlated with) the actions taken by the students. The features were emergent from our coding process, with the goal of defining features that could be unambiguously assigned to each project and were common to at least two projects. We describe both the modeling and feature codes further in Sec. IV.

Upon completing the code book, we conducted a confirmatory interrater reliability (IRR) check. This consisted of checking all of the code instances appearing in two of the projects for false positives. Since there is no natural unit of coding given the varied data sources, a single code instance could be any length of connected text or part of a figure or presentation slide. Initially, the coders had 100% agreement on the feature code instances and 93% agreement on the modeling code instances, and they reached 100% agreement on both after discussion. This IRR process led to the removal of one feature code and minor revisions and clarifications to several other code definitions.

For the analysis, we chose the projects (instead of the students) as the unit of analysis and chose to look at the presence or absence of codes within a project instead of the total number of instances of each code. This allowed us to investigate whether or not the projects afforded opportunities for modeling, especially when answering RQ2, independent of whether one or all of the students working on that project engaged in modeling in a specific way. During the IRR, however, we checked by instance to confirm the coding process.

D. Limitations of study

There are three primary limitations of this study: the uneven data sources for the different projects, the limits to the generalizability of a small-sample-size study, and the onset of the COVID-19 pandemic during data collection. The different data sources available for each project provided us slightly different information, both the amount and the kind, about each project. For example, some of the lab notebooks contained details about all the actions the students took without specifying their underlying reasoning, whereas the interviews provided in-depth explanations about only a few decisions the students made. Because of the potential for missing information, this analysis may be an underestimate of the amount of modeling done in each project. We minimize this effect by analyzing the data by project instead of by number of code instances within each project.

Additionally, the sample of students in our study may limit the extent to which we can generalize the results

presented in this paper [51,52]. Our data come from 36 students working on 14 different projects at three distinct kinds of institutions, so we do not endeavor to generalize our conclusions to all undergraduate physics students. Our primary intention is to identify specific ways students *might* engage in modeling in multiweek projects, and not to make generalizable claims about the amount of modeling we would expect from students for all advanced lab projects.

Although this research project was not intended to study remote courses, the onset of the COVID-19 pandemic affected our results in a variety of ways. Many of the projects were designed knowing they had to be performed at home, which provides a different set of projects to analyze than what would be found in a typical term. Some of the students from course 3 had to suddenly change their projects in the middle of working on them, eliminating the opportunity for the extensive planning process normally provided in the course. Many of those students did not write in their lab notebooks, affecting both the students' experiences of authentic scientific practices and the amount of data available for analysis. Additionally, we may be missing out on the voices of students who were most affected at the start of the pandemic because they may not have been able to engage as fully in their courses or participate in the interviews.

IV. PROJECTS, FEATURES, AND MODELING CODES

The first outcome of our coding analysis is an understanding of the different projects, the features we attribute to the projects, and the elements of the EMF with which the students engaged. The 14 student projects included in this analysis are summarized in Table III. We include information about the project goal, the course, whether or not the project was completed remotely, and the features assigned to each project. Some of the students changed their goals while working on the projects, so we include the final goal as appearing in the final reports or presentations and note that not all of the projects achieved their goals. The rest of this section describes the feature and modeling codes.

A. Feature codes

The feature codes are independent of the EMF and are emergent from our coding process. They come in sets (labeled by letters), where each project was assigned one feature within each set. Here is a list of the features with identifiers:

- A1. Build complex apparatus
- A2. Build intermediate-complexity apparatus
- A3. Build simple apparatus
- A4. No building

TABLE III. List of the 14 student-designed projects. The names are used as project identifiers in Figs. 3 and 5. The modality denotes whether all the students on each project worked remotely, went in to the lab at the same time, or went into the lab at different times from one another in order to satisfy social distancing protocols. The feature labels are defined in Sec. IV A.

Name	Project goal	Course	Modality	Features
Acoustic levitation	Image acoustically levitated water and compare with normal 2D imaging	3	All remote	A1, B1, C2, D1, E1
Blackbody radiators	Measure emissivity of light bulbs, relationship between power and temperature, and the Stefan-Boltzmann constant	1	All remote	A3, B1, D2, E1
Chain fountain	Investigate dynamics of chain fountains and the dependence on chain density	1	All remote	A3, B2, C1, D3, E1
Earth's magnetic field	Compare three methods of measuring earth's magnetic field	1	All remote	A3, B2, C1, D1, E1
IBM quantum computer	Compare measurement of entangled vs nonentangled probes on IBM quantum computer	3	All remote	A4, C1, D2, E2
Model rockets	Measure the coefficient of drag of model rockets	1	All remote	A2, B2, C1, D3, E2
Optical pumping	Use optical pumping to measure the cross section of absorption and the Landè g factor	2	In person, together	A1, B1, C2, D2, E1
Paschen curve	Measure relationship between breakdown voltage and pressure of a plasma (Paschen's curve)	3	All remote	A4, C2, D2, E1
Phonons	Measure phonon scattering spectrum of quartz	3	In person, separately	A1, B1, D2, E1
Photon detector	Build and characterize a single photon detector	2	In person, together	A2, B1, C1, D1, E1
Refractive index	Measure relationship between refractive index and temperature for different oils	1	All remote	A3, B2, C1, D3, E1
Sonoluminescence	Measure relationship between current and intensity of sonoluminescence	3	In person, separately	A1, B1, C1, D3, E1
Tensile strength yarn	Measure tensile strength of 2-ply wool yarn	1	All remote	A2, B2, C1, D3, E1
Van der Pol oscillator	Build and model a Van der Pol oscillator	1	All remote	A2, B1, C1, D3, E2

- B1. Use only lab equipment
- B2. Use some equipment from home
- C1. Conceive and plan at least part of apparatus
- C2. Don't conceive and plan apparatus
- D1. Goal: Characterize apparatus
- D2. Goal: Question with known solution
- D3. Goal: Question with specific instance unknown
- E1. Established model exists
- E2. Requires model construction

Several of the features relate to the students' building of apparatus: whether or not they built an apparatus themselves, how complicated the apparatus was, and what kind of equipment was needed for building. Features A1 to A4 specify the complexity of the apparatus the students built or if they did not build one at all. Different students may find different aspects of the experiments complex, and this may be different than what the instructor would find complex. Thus, we define complexity in terms of the number of parts required to build an apparatus and how much time an expert would need to spend setting it up once the parts were assembled. The precise definitions for the differences between complexity of apparatus, and all of the other feature codes, are given in Table V in Appendix B. In our dataset, the two projects that did not build anything both remotely operated a publicly available apparatus, one of IBM's quantum computers or an experiment at the Princeton Plasma Physics Laboratory. The other features related to building are B1 and B2, which distinguish between projects where the students built an apparatus entirely with lab equipment and ones where they used some materials more commonly found in a home (e.g., plastic containers and cell phone sensors). Note that the students who operated an apparatus remotely are not coded in either of these categories, so the feature A4 additionally completes this feature set.

Features C1 and C2 provide information about whether or not the students conceived of the idea for their apparatus and procedure themselves. It is possible for students to plan an experiment without building it or to build an experiment without planning it. For example, some of the students built an apparatus by following a manual instead of planning it out themselves. Conversely, the students working on one of the projects that did not build anything had a lot of control over the procedure of the remotely operated apparatus and needed to plan out which parts of the apparatus would be used for their experiment. For two of the projects, we were not able to determine whether or not conceiving and planning the apparatus was part of the projects, so they are not assigned either of these features.

The remaining features relate to the goal of the project: what kind of goal the project had and whether or not model construction was required to achieve that goal. We assigned three possible goal features. The feature D1 was assigned to projects that aimed at gaining a better understanding of an apparatus (or several apparatus) and how well it worked for

a specific application. The feature D2 was for projects whose goal was to answer a question where the solution was known to the scientific community. The feature D3 was for projects whose goal was to answer a question to which the students believed the specific case they were investigating was unknown, although results for similar systems may have been well known. The last two features (E1 and E2) indicate whether or not model construction was a required component of the project based on the project goal. Model construction could have been a necessary step for either figuring out how to build the apparatus or creating a prediction with which to compare the experimental result. Note that feature E2 was assigned not when we saw evidence of students constructing a model, but only when there was evidence that the project goal could not be accomplished without model construction.

B. Modeling codes

The modeling codes we use to demonstrate the elements of the EMF with which students engage consist of both *a priori* codes representing the main tasks of the EMF and emergent subcodes for some of those tasks. A diagram of these codes is shown in Fig. 2 with the main modeling tasks in the left column and the emergent subcodes in the right column. We developed subcodes for the *engage with models* and *revisions* tasks because those actions had the most variation across projects. Full definitions of all the modeling codes are provided in Tables VI–VIII in Appendix C.

The definitions for the main modeling task codes follow almost directly from the EMF. The code *make measurement* is assigned any time the students make a measurement. We label the next tasks *engage with models* to account for the variety of ways the students engage with models without necessarily constructing them, retaining the distinction between the measurement system and the physical system. The code *analyze data* consists of all steps the students take to convert the data into a form that permits making a comparison with a theoretical model prediction. The codes *make comparison* and *propose causes* are assigned any time the students make a comparison between an experimental result and a theoretical prediction and suggest possible reasons for any deviation between the two. Both of these codes can be used either for a comparison of the project goal or for a troubleshooting comparison that may be either qualitative or quantitative. Lastly, the students may choose to revise the apparatus or model of either the physical or measurement systems.

Although there may be some ambiguity over what counts as part of the physical versus the measurement system, we distinguish between the two systems based on the question the experiment is trying to answer. For some experiments, there is a clear division between the physical and the measurement systems. For others, the two parts may be more integrated, leading to multiple valid ways to classify

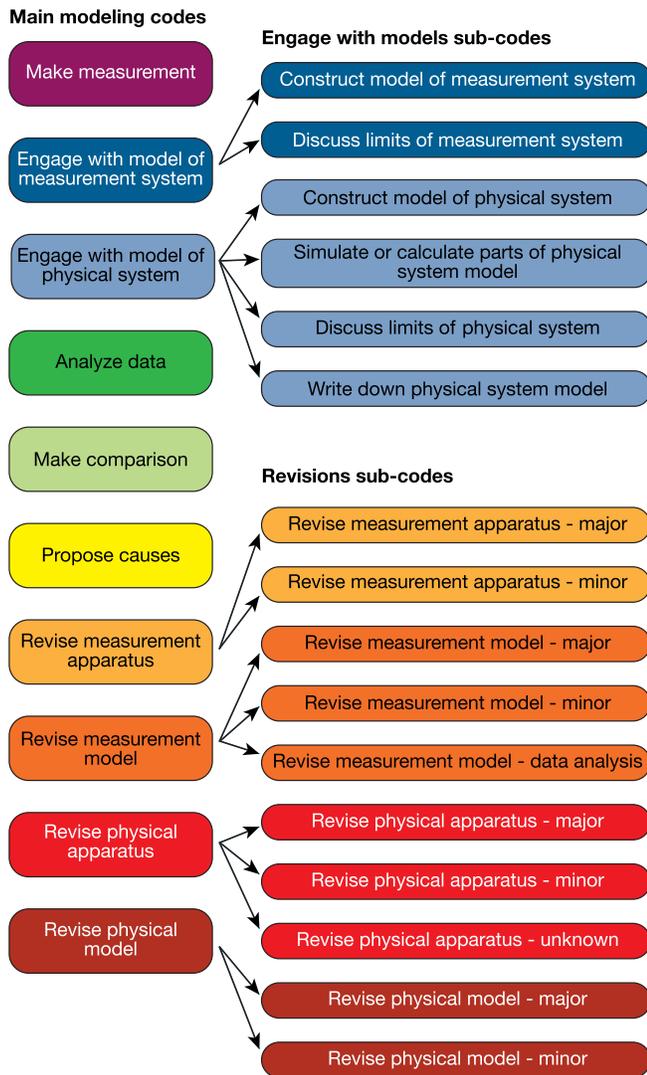


FIG. 2. Modeling codes. The main modeling codes coming directly from the EMF appear on the left, and the arrows point to subcodes for the *engage with models* and *revisions* codes.

the parts of the experiment [4]. For this work, we define the physical system as all parts of the experiment that are needed to create and model the system the experimental question is about. Conversely, the measurement system is all parts of the experiment needed not to define the experimental question but only to answer it.

The *engage with models* subcodes (the first set in the right column of Fig. 2) describe student actions related to model construction ranging from finding a model in the literature to constructing a model themselves. Some of these subcodes come directly from the EMF (for example, limitations and assumptions of the models is listed as part of model construction in Fig. 1), while the other subcodes are emergent from actions the students took, as documented in the course artifacts. The two construct-model subcodes are assigned to projects where the students created a model themselves as opposed to applying a model they found in

the literature. Another action students took was using either calculations or simulations to apply an already existing model to their specific experiment. We did not see clear indications of analogous behavior for applying models of the measurement system, although some similar actions were coded as *data analysis* because the students did not explicitly mention models of the measurement apparatus. Students also discussed limitations or assumptions of their systems or models. The code *write down physical system model* was given when the students described a general model of their physical system as it would appear in an article or textbook, without necessarily relating it to their own experimental parameters.

Revisions in the EMF are already divided into four categories based on what is being revised, and we further subdivide them into major and minor revisions. Major revisions involve adding, removing, or switching out one part of the apparatus or model, for example by replacing a component in a circuit or using an entirely new measurement procedure. Minor revisions involve only slightly altering the already existing apparatus or model, such as by fixing a minor mathematical mistake or realigning a mirror. We chose to code this distinction because the extent of the revisions the students performed fell on a spectrum, and we believe there is a fundamental difference between the two extremes (e.g., catching a sign error versus constructing an entirely new model). Note that revising the measurement procedure is coded as a part of revising the measurement model, and can be either minor or major, depending on the extent of the revision. We classify a revision of the data analysis method to be its own subcode of measurement system model revisions, because it is a common action that is often performed without explicit consideration of the model of the measurement apparatus. There is one additional subcode, a revision of the physical system apparatus of unknown scale, with a single instance where the students discussed a revision without giving enough details about it for us to know whether to classify it as major or minor.

V. TRENDS ACROSS PROJECTS

In this section, we present the number of projects assigned each modeling code. Our intention is not to focus on the specific numbers, but to look at overall trends in order to understand which parts of the modeling process were performed by students during their projects and how that differs by project. In Sec. VA, we answer RQ1, how students engage in the modeling process, by providing the number of projects in which students performed each of the tasks of the EMF. In Sec. VB, we answer RQ2 and describe associations between the features of the projects and the different ways the students engaged with model construction and revisions. We focus on model construction and revisions because students exhibited a large variation in ways they engaged with those tasks, and the tasks are goals

of the instructors of the three courses. In-depth examples and discussions expanding on these results are provided in Sec. VI.

A. Student engagement in modeling

To investigate how students in our study engaged in modeling (RQ1), we first determine which projects displayed evidence of each of the main tasks of the EMF. These results are shown in Fig. 3, where we see some differences among the projects in which modeling tasks were performed. Nonetheless, the majority of the boxes are “Yes” showing that the students overall performed many of the modeling tasks described by the EMF.

The modeling codes *engage with model of physical system* and *make comparison* were performed by students in all of the projects, possibly due in part to the project guidelines. As described in the rubrics for the final reports or presentations, the students were graded on whether or not they explained the physical system model and made a comparison. Because the code *make comparison* includes both qualitative troubleshooting comparisons and comparisons of the goal measurement, it was possible for projects to be assigned this code even if the students were not able to make a measurement of their goal quantity, as was the case for the Acoustic Levitation project.

Additionally, the students in all but one of the projects performed the tasks *make measurement* and *propose causes*. This is consistent with prior research showing that students usually make measurements, yet they are less likely to propose causes unless prompted to do so [43,47]. In our study, we assigned the *propose causes* code to projects that proposed causes both in the middle of the modeling process (e.g., as appearing in lab notebooks) and

after the students were prompted to reflect on possible sources of error at the end of the process (e.g., as appearing in final reports and presentations).

The least commonly done tasks were *revise physical model*, *revise measurement model*, and *engage with model of measurement system*. Many of the projects had established models with which to compare their experimental results, so the students did not need to engage in model construction or model revisions in order to achieve their goal. Furthermore, many of the students used common measurement devices (e.g., oscilloscopes, voltmeters, and cameras) without considering how they functioned, so engaging with or revising a model of the measurement system occurred less frequently than with the physical system model.

To better understand the differences between projects, we next consider the prevalence of the *engage with models* subcodes. The details of this analysis are shown in Fig. 5(a) in Appendix D and summarized here. In all of the projects, students wrote down a model of the physical system, but they did not necessarily relate that model to their specific experimental parameters. Full construction of a model of either the measurement or physical systems occurred in fewer than half of the projects. However, students in many of the other projects still engaged with models in other ways, such as by performing simulations or calculations or considering limitations and assumptions of one or both of the models.

There was also a variation across projects in the kinds of revisions that were performed, with each project being assigned between zero and five of the *revisions* subcodes. Many of the apparatus revision subcodes were the most prevalent, with major revisions to the measurement apparatus being the most common type of revisions, followed by

	Acoustic Levitation	Blackbody Radiators	Chain Fountain	Earth's Magnetic Field	IBM Quantum Computer	Model Rockets	Optical Pumping	Paschen Curve	Phonons	Photon Detector	Refractive Index	Sonoluminescence	Tensile Strength Yarn	Van der Pol Oscillator	N
Make measurement	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	13
Engage with model of measurement system	No	No	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	No	Yes	No	7
Engage with model of physical system	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	14
Analyze data	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	11
Make comparison	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	14
Propose causes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	13
Revise measurement apparatus	No	Yes	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	10
Revise measurement model	No	No	Yes	Yes	No	Yes	No	No	No	No	Yes	No	Yes	Yes	6
Revise physical apparatus	Yes	No	Yes	No	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes	10
Revise physical model	No	No	Yes	No	No	Yes	No	No	No	No	Yes	No	No	No	3

FIG. 3. Presence (“Yes,” blue) or absence (“No,” white) of the main modeling codes in each of the 14 projects. The number of projects that demonstrated evidence of each task are listed in the far-right column.

minor and major revisions to the physical system apparatus. Most major revisions to the measurement apparatus occurred when students switched to using an entirely different measurement device (e.g., switching from an oscilloscope to a multimeter), with the hope it would be more accurate or work better. The model revisions, aside from revisions to the data analysis method, were the least common with only one or two projects each being assigned the codes major and minor revisions to the measurement and physical system models. More details are provided in Fig. 5(b) in Appendix D.

B. Associations between project features and modeling

Our second research question investigates connections between the project features and the ways students engaged in modeling in the projects with those features. More specifically, we again look at the occurrence of the *engage with models* and *revisions* subcodes but by project feature instead of by project. The goal of this analysis is to better understand the kinds of projects that support different forms of student engagement with models and revisions, leading to potential implications for instruction.

Figure 4 shows the co-occurrence of features and *engage with models* subcodes for all the projects in this dataset. We omit the subcode *write down physical system model* since the students in all of the projects performed this action, so it does not help us understand the variation in student engagement in modeling. The shading represents the fraction of projects with a given feature (columns) with evidence of engagement in the specific type of activity (rows). The column averages are shown at the top of the plot, and they indicate which features, on average, have a

higher prevalence in projects with evidence of engagement with models. We note that there is no reason to believe the students in a single project should perform all of these actions in order to be successful; instructors may have specific learning goals related to only some of the subcodes, and thus the column average may not be an important metric for certain contexts.

As shown in Fig. 4, there is a dependence on project feature for many of the *engage with models* subcodes. The features with the highest column averages are more strongly associated with model engagement for the projects in this dataset. These associations are most evident for the modeling tasks of constructing models of the physical and measurement systems, whereas, the subcode *discuss limits of physical system* appears at similar rates across all project features. Table IV lists the six project features with the strongest associations with model engagement on average. Although there is no obvious metric for what counts as a “strong” association, we include the feature with the largest column average from each feature category (or two in the case of a tie), which give us the six features with the highest column averages overall. Additionally, these features appear as possible causes for the specific instances of modeling discussed in Sec. VI.

We perform an additional analysis of associations between revisions and project features. We find associations between project features and model revisions that are similar to those between project features and engagement with models because major model revisions are often instances of model construction. We also find differences across project features for measurement system revisions, which are dominated by the differences in revisions of the

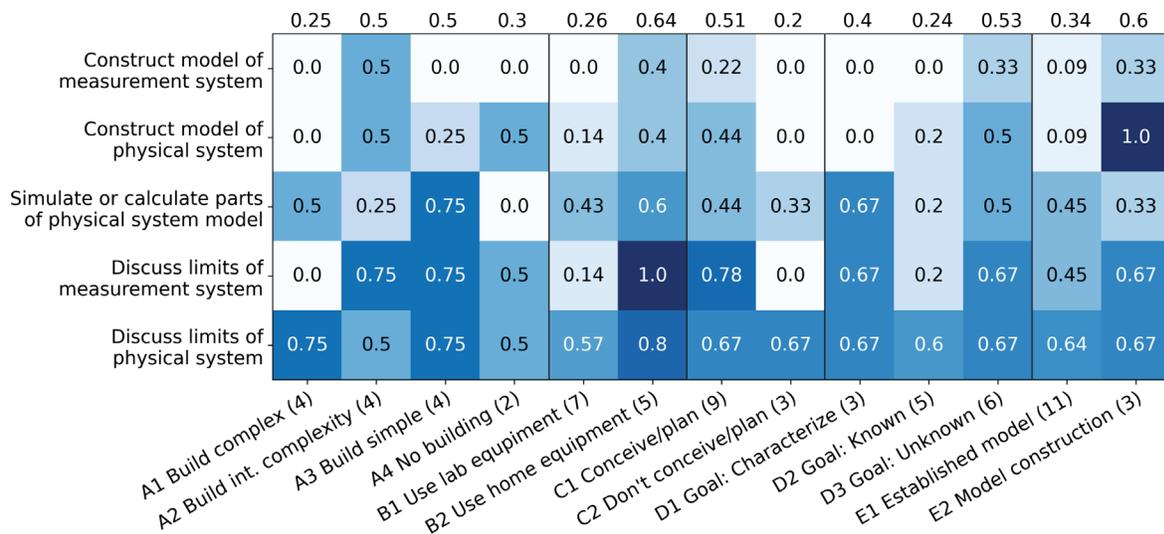


FIG. 4. Fraction of projects with a specified feature (columns, organized by category) assigned the *engage with models* subcodes (rows). The numbers in parentheses following the abbreviated feature names are the number of projects with each feature, and the numbers on top are averages over each column. The higher the column average, the more the feature may be associated with engagement with models on average. The modeling codes are sorted by average over row, so the lower modeling codes more frequently appeared in student projects.

TABLE IV. Project feature codes most strongly associated with model construction along with their shortened definitions and example quotes from the student data sources (full definitions are in Table V in Appendix B).

Feature code	Definition	Example
Build intermediate-complexity apparatus (A2)	The project provides the students the opportunity to build at least part of an apparatus that has approximately 5–10 components and requires some effort to assemble. It does not require careful alignment, so once it is fully put together it should work easily.	“The circuit consists of... an op-amp comparator... The signal entering the positive terminal... is generated by the... LED in series with a... variable resistor and the signal entering the negative terminal... is generated by a... potentiometer in series with a 5V source.”
Build simple apparatus (A3)	The project provides the students the opportunity to build at least part of an apparatus that involves only a few parts. It should be able to be assembled by an expert in a few minutes with a high probability of working immediately.	“...[the set ups] had aimed a ... red laser pointer on an oblique Pyrex container containing the sample material (oils of various kinds) as temperature was varied.”
Use some equipment from home (B2)	Some of the equipment required for the project is more commonly found in a standard home than in a typical physics lab. Examples include plastic containers and sensors on a phone.	“Model rockets can be quite easy to build at home; in fact, a simple construction with some plastic pipe, vinegar, and baking soda is possible.”
Conceive and plan at least part of apparatus (C1)	The project provides the students the opportunity to come up with ideas for at least part of the apparatus and/or procedure themselves.	“...we tried to develop our own little experiment... we weren’t necessarily basing it off of specific processes or setups that we had seen in research...”
Goal: Question with specific instance unknown (D3)	The primary goal of the project is to measure some property that is known for other similar apparatus or other materials but students believe is not known for the specific case they are investigating.	“...we experimentally find the relationship between the tensile strength and the twist angle of two-ply wool yarn. We then compare our results with Huang <i>et al.</i> ’s findings on two-ply polyester yarn...”
Requires model construction (E2)	The project requires the students to do the majority of the model construction on their own in order to build the apparatus or have a theoretical prediction with which to compare their experimental results. There is no single place they can look up a model in the literature.	“... we spent the most time on... that link between physical system apparatus and model construction... Cause the Van der Pol equation is what we were interested in and it’s... a pure math object... and this circuit... describe[s]... the same oscillation.”

measurement models. There is not as much feature dependence of the apparatus or physical system revisions, possibly because the students in most of the projects revise the physical system apparatus in some way. More details are provided in Appendix E.

VI. INSTANTIATIONS OF MODELING

In this section, we expand on the results presented in Sec. V and provide concrete examples of student engagement in modeling. These examples consist of student quotes from the reflections, interviews, and responses to audience questions after the final presentation. We first present specific instances of model construction in the projects demonstrating the different paths students took to constructing models. Each of these paths is connected to some of the project features most associated with model construction as shown in Fig. 4 and Table IV. We then discuss a potential trade-off between students working with complex equipment and constructing models and enacting revisions to the measurement system.

A. Paths to model construction

In the projects in our dataset, we identify three paths to constructing models: (1) students constructed a model of the physical system when such a model was required in order to build the apparatus, (2) students constructed (or attempted to construct) a model of the physical system in order to compare an experimental result with a prediction, and (3) students constructed a model of the measurement system when they had to conceive of a way to make the desired measurement. We further discuss which of the features most associated with model construction may lead to students embarking on each of these paths.

1. Model construction of physical system to build apparatus

The first way we saw modeling occur was when it was required in order for the students to set up the physical system apparatus. This appeared in two projects, the Van der Pol Oscillator and the IBM Quantum Computer. In both, the students needed to use general knowledge

about electronics and quantum states to construct a model in order to set up the experiment, whether or not they did the building themselves. In the final student reflection, when asked how he had grown due to his work in this course, one student from the Van der Pol Oscillator project discussed how his understanding of what modeling could be used for changed from only answering theoretical questions that could not be answered experimentally to being necessary for developing an experiment. In particular, he mentioned how his group had to figure out which electronic parts were needed to build the desired circuit:

“...a computer model was crucial to the development of our experiment, as we needed the full in-depth model in order to identify the proper components (capacitors, inductors, measurement devices) that we would need to obtain observable oscillations from our circuit.”

Note that the student is talking about components that we consider to be parts of both the physical and measurement systems.

Both of the projects that needed to construct models in order to set up their experiment were coded with the feature *requires model construction*. This is because there was no single model they could take from the literature that matched perfectly with what they wanted to do. The projects also share the feature *conceive and plan at least part of apparatus* because model construction was part of planning their apparatus. When model construction is required in this way, the other features have less of an effect over whether or not the students construct a model. For example, the remaining features in these two projects do not overlap. One of these projects had a goal with a known solution and the other had a goal with an unknown solution. One of them did not even build anything. Independent of the other project features, if a project requires model construction in order to initially set up the apparatus, the students must construct a model to even begin the experimental process.

2. Model construction of physical system to make a comparison

The second way model construction occurred was when the students did not have an already-existing sufficient model with which to compare their experimental results. The comparison stage is a key part of the EMF and was a requirement for these course projects. One of the projects that constructed a model for comparison was Model Rockets, which consisted of building simple rockets that were easy to build and launch at home, but complicated to fully model. When asked in the reflection to identify a challenge they faced in the past week, one of the students described their attempts at modeling:

“One of the major problems we ran into was theoretical modeling. At first, we wanted to directly model the

chemical and physical state of the water rocket. Eventually we realized the models and simulation would have to be absurdly sophisticated, and even then, we could easily make a mistake or neglect a small effect which the system is very sensitive to, ruining our predictive power. Therefore, we decided to change up the flavor of our project, shifting more to a measurement-based experiment rather than a theoretical numerical project.”

These students switched from constructing a model of the chemical process to using a simpler model with forces and measured accelerations.

The other project that attempted to construct a model to better understand their experimental data was the Chain Fountain project. Those students initially compared their results with a model from a research article and realized it was not sufficient to explain the differences they saw in their experimental results for chains of different densities. In response to a reflection prompt asking about challenges they faced, one of the students said

“The biggest challenge I encountered was trying to figure out how to model our system. We were not sure how to calculate the differences between bead types or pot angle. Sadly, we never really solved this problem.”

In this project, the students did not need to construct a model to make a measurement or even to make the initial comparison with a prediction. However, after performing the measurement, the students realized the model they were using did not account for some factors such as bead type or pot angle, so they tried to construct a model on their own. They ultimately were unsuccessful, but we suspect this exercise in model construction could still have contributed to their development of modeling skills.

One of the features these projects share that could be contributing to the need for model construction is *goal: Question with specific instance unknown*. In order for the students to need to construct a model or adapt a model to fit their exact project, the existing models must not be sufficient, even if the students do not know this initially. Thus, the students must not already know the answer to the question they are asking, because if a good model existed, it would already accurately predict their experimental results. This is distinct from the feature *requires model construction* because the students may be able to make a comparison without constructing a model, as is the case for the Chain Fountain project. There, the students compared their measurements with a model found in a research article, and the comparison showed the model was inadequate. It is interesting to note that both of these examples are of failed attempts at model construction. When model construction is needed only for the comparison stage or after, the students are able to engage in a significant portion of the modeling process whether or not the model construction is successful.

In contrast, one of the students working on a project with a known solution and an established model acknowledged that his lack of engagement in modeling, as depicted by the EMF, may be due to asking the wrong question. When asked in an interview what his group did when they got to the comparison stage of the framework, this student said

“But it wasn’t a very good question, because we already knew what the answer was going to be.”

When asked if he would do something differently if he were to do it again, he responded

“I wanted to do something like looking at maybe explore questions that aren’t... comparing ... what we found to what someone else found, or maybe do more qualitative questions, like, does the curve stay consistent for different gases...?”

This student was working on the Paschen Curve project, which, due to the COVID-19 pandemic, switched to remotely operating an apparatus where the students had very little control over the questions they asked. They ended up taking a measurement and then stopping because it matched the theoretical prediction. For them, asking a question with a known solution meant they did not have a reason to construct a model or iterate on their experiment.

3. Model construction of measurement system to make a measurement

The third way model construction occurred in our dataset was when students needed to build their own measurement device. In order to know which parts were required for measuring their desired quantity, the students had to construct a model of how their intended device worked. One of the students from the Tensile Strength Yarn project discussed figuring out how to measure the tensile strength of the yarn, saying how it is done commercially with a specific machine designed for that purpose, and they had to conceive of a way to measure it at home. In response to a reflection prompt asking about the biggest challenge they encountered on the final project, another student in the group said

“Making sure the measurement procedure was consistent was also a challenge as I was indirectly measuring force by measuring mass and we didn’t have a force meter or could [sic] conceive of a way to properly use one for what we wanted to test.”

The lack of an already-existing measurement device led to this group modeling and then building their own. Both of the projects that constructed models of the measurement systems were ones with the features *use some equipment from home* and *conceive and plan at least part of*

apparatus. The feature *use some equipment from home* may have been present because there are fewer commonly used methods to make complicated measurements at home, so students needed to design (and therefore model) their own.

B. Potential trade-offs between modeling and complex equipment

One theme we see in the data is a potential trade-off between the students working with complex equipment typical of research labs and engaging in modeling practices. This is evident both in the overall trends of associations between features and modeling discussed in Sec. VB and directly in some student quotes. Projects with complex apparatus easily allow for the students to spend most of their time building a functioning apparatus instead of focusing on model construction or revisions to the measurement process.

One of the students identified this trade-off when discussing two possible projects, one with a theoretical focus and the other that involved working with technical hardware. The student had planned to work on a single-photon interference experiment, but had to switch to operating one of the IBM quantum computers due to the COVID-19 pandemic. In response to a question asked after his final presentation about comparing what he would have learned from the two projects, he said:

“[For our initial project idea], I think I had a pretty good understanding of the theoretical backing and like how to implement it. So the real challenge, and what I would have learned was about doing hardware, right? Like how do I run a single photon detector and things like that... And this one, it was the other way around. So we didn’t really have to deal with hardware very much. Like our only implementation was writing the coding, but figuring out the theory is like how we are going to build these circuits, figuring out, like, how we can reconstruct the angles, given this information. I learned a lot more about that side of it from this project than I would have from that one.”

The student discusses how he ended up doing a lot of theory (which includes model construction), but he would have spent much more time learning about hardware instead if his plans had not been disrupted.

We see this potential trade-off between constructing models and gaining experience working with complicated equipment in the other projects in our dataset as well. All of the projects that fully constructed models (aside from the one that remotely operated an apparatus) involved simple or intermediate-complexity apparatus; none of them built a complex apparatus. All of the projects where the students built a complex apparatus had an already established model with which to compare. The students working on those

projects spent a lot of time attempting to precisely set up the apparatus and therefore may not have had time to focus on the models. Our evidence is consistent with a study investigating in-person student-designed multiweek projects that found that students who chose projects with complicated apparatus spent most of their time assembling or adjusting the apparatus instead of engaging with models [53].

Another place we see this potential trade-off is with revisions to the measurement system. Many of the projects with the feature *build complex apparatus* spent most of their time constructing the physical apparatus and several did not have a chance to iterate on the measurement system because they did not take an initial measurement of their goal quantity. For example, the projects Acoustic Levitation and Sonoluminescence were never able to see their desired signals, even by eye, so there was no reason for them to focus on the measurement device. In contrast, many of the projects that built simple and intermediate-complexity apparatus enacted different kinds of measurement system revisions.

The projects with simpler apparatus allowed the students to make a measurement earlier in the experimental process, leaving time for revisions to the measurement system. For example, the Chain Fountain project, whose physical system apparatus consisted of a chain flowing out of a container, spent most of their time changing parameters of the videos they took, figuring out better ways to track the chain as it fell, and then revising their method of analyzing the camera frames used to measure the velocity. Another example is the Refractive Index project whose apparatus consisted of oil, a container, and a laser pointer. One of the students working on that project discussed their plans to iterate on the measurement procedure in a reflection prompt asking about one of their project goals for the following week:

“To [take our measurements], we have split up responsibilities for taking our ‘rough’ first measurements to ascertain issues and strengths of the procedure in trial runs, and then after discussing how each attempt went, we will reform the procedure and repeat measurements more formally.”

Because their physical system apparatus was simple, they were able to spend most of their time revising the measurement system.

VII. IMPLICATIONS FOR INSTRUCTION

Our analysis demonstrates that in the courses we investigated, certain project features were associated with student engagement with modeling and revisions, leading to the possibility for instructors to guide students towards projects with features aligned with their learning goals. Instructors may have a variety of different learning goals

related to student-designed projects; for example, they could want students to fully construct models, consider limitations of measurement system models, iterate on one specific aspect of the project, or learn how to work with a specific piece of technical lab equipment. While acknowledging there is a wide array of possible desired course outcomes, here we summarize our findings from Secs. V and VI into three main takeaways for instructors.

The features of projects chosen can affect how the students engage with models, including both the need for, and the success of, model construction. For the courses we studied, we found that the features listed in Table IV appeared more frequently in projects in which the students constructed models, with different features being connected to different ways the students constructed models. If an instructor wants model construction to be a necessary part of the experiment, they can help students devise projects that require model construction either in order to build an apparatus or to make a comparison. However, whether or not model construction is feasible may not be clear at the question-defining stage. Instructors may also guide the students towards research questions where the answer is not fully known and there is some ambiguity over whether the model the students compare with is valid for their experiment. In this case, the students may not successfully create a model, but the process may still hold value. One other path to model construction is when there is no obvious method of making the desired measurement, which depends both on the research question and the available equipment.

Instructors may similarly guide students towards projects with specific features if they desire the students to enact certain kinds of revisions. In particular, an instructor may suggest students use a simple or intermediate-complexity apparatus if the goal is for the students to have time to focus on measurement system revisions. As long as the students build at least part of an apparatus, they will probably enact physical apparatus revisions. However, not all possible projects will lead to the students performing either measurement system or model revisions. Since some model revisions are a form of model construction, the features associated with model construction may also lead to model revisions. In the projects we analyzed, we saw that similar features were also associated with the students enacting revisions of the measurement system, although many projects, even without these features, swapped out one measurement device for another. The students that did not need to spend time aligning complex apparatus had more opportunities to iterate on other parts of the measurement system.

Students are able to engage in modeling without expensive equipment. All of the projects within this dataset that constructed models were done remotely—using equipment from home, remotely operating a publicly available apparatus, or using only inexpensive lab equipment. In some of these projects, there was a trade-off where the students did not gain experience working with equipment

commonly found in research labs, but they did have the opportunity to practice model construction. If an instructor's goal is to focus on modeling, it can be done with relatively simple and cheap equipment.

In order for any of these ideas to be implemented, instructors must be able to help students choose or refine their project topics. This guidance can take many forms. In all three courses analyzed, the instructors provided the students with a list of example project ideas and had a role in helping the students choose their final topics. The exact method of guidance chosen would depend on instructor preference and local course context. Possibilities include carefully considering the features of listed example projects, explicitly requiring the students to pick a project with a certain set of features, and helping the students refine project ideas and goals through one-on-one meetings or feedback on project proposals.

VIII. CONCLUSION AND FUTURE DIRECTIONS

We have analyzed student course artifacts and interviews from three different upper-division lab courses with multi-week student-designed projects to better understand how students engage in modeling within a class setting. Using the EMF as an analytic tool, we answered both of our research questions. For RQ1, we found that most students engaged with many different parts of the modeling process. However, there was a large variation across projects in whether or not students constructed models themselves and on which parts of their experiment they iterated if they did not achieve the desired result. For RQ2, we found that students who worked on projects with certain features (see Table IV) were more likely to construct and revise models and revise the measurement system. Other parts of the modeling process, such as making measurements and revising the physical system apparatus were done by most students independent of the features of their projects.

Although this work was not intended to investigate student-designed projects done at home or remotely, the timing of the COVID-19 pandemic provided us a unique dataset from which we were able to analyze projects with certain features (e.g., projects that used some equipment from home or that used simple or intermediate-complexity apparatus) that we may not otherwise have had access to. While this could be seen as a limitation of this work, we instead see it as an opportunity to realize the potential learning opportunities afforded from working with low-cost equipment. Even without expensive lab equipment, several of the projects in our dataset provided the students the opportunity to practice modeling skills. Different project features led students to engage in different aspects of modeling, and this work contributes to the understanding of how to best align student-designed projects with course learning objectives.

Future work could extend this type of descriptive analysis to different contexts to improve our understanding of ways to

engage students with modeling. Applying a similar analysis to a different set of courses, particularly in-person ones, would help instructors generalize results more broadly. Additionally, using the EMF as a tool to analyze how undergraduate students engage in modeling while participating in research could help extend prior research [11] to understand which aspects of undergraduate research experiences these student-designed projects in lab courses could reproduce or improve upon. This study investigates students in advanced courses and assumes the students are already capable of undergoing the modeling process. In order to prepare students for the learning possibilities offered by these open-ended projects, instructors may need to scaffold model construction and iteration in earlier courses [47]. However, how to best provide scaffolding without removing student agency is still an open question.

The analysis presented in this paper focuses on how project features relate to student engagement in modeling, but we could additionally investigate the interplay between other factors, such as students' views of experimental physics or group dynamics, and engagement in modeling. Further work could connect students' epistemological views with the amount and kind of modeling the students undertake and examine how those views change due to engagement with models during self-designed projects. In our dataset, different groups of students divided up the tasks in various ways, leading to some students engaging with all aspects of their experiments, while others focused solely on taking measurements or running a simulation. Although we were not able to investigate these differences with the available data sources, future work could investigate if the division of work led to any differential learning among group members and how group dynamics affect student engagement in modeling.

ACKNOWLEDGMENTS

We thank the instructors and students of the three courses for partnering with us and sharing their experiences and coursework, Dimitri Dounas-Frazer and Laura Ríos for initialization of ideas and data collection that preceded this work, the CU PER group for useful conversations and feedback, and Alexandra Werth and Michael Vignal for feedback on the manuscript. This work is supported by NSF Grants No. DUE-1726045, No. PHY-1734006, and NSF QLCI Award OMA 2016244.

APPENDIX A: DETAILS OF METHODOLOGY

This section contains additional details of our methodology.

1. Courses

a. Course 1

Course 1 is an advanced lab course often taken by junior or senior physics majors with the primary goal of having

them learn about different aspects of being an experimental physicist. This course is taught at a small, private, predominantly white, liberal arts college in the United States. As described in the course syllabus, the course has four main learning objectives: the students will learn to clearly communicate experimental results, be reflective about experimental physics, evaluate the quality of data and compare it with an already existing model, and collaborate on and record the process of making a measurement and refining the methodology. In the spring of 2020, there were 24 students enrolled in this course, and they worked in instructor-assigned groups of three students for the course activities. The entirety of the course was taught remotely due to the COVID-19 pandemic, and the instructor chose to retain all of the course objectives, while adapting the final project to be conducted from home.

This course consisted of two two-week-long structured labs prior to the four-week-long student-designed final project. Due to needing to suddenly adjust the course for the remote format, the course was altered slightly from how it had been taught in previous years, and the usual 10-week-long term was shorted to nine weeks. In one structured lab, the students characterized properties of different metals. They were provided with data taken by the instructors and were tasked with focusing on data analysis and the presentation of the results. The other lab was focused on signal processing. The students characterized a voltage amplifier by remotely operating lab equipment set up by the instructors. After completing each lab, the students were required to write a lab report, one of which was written as a group and the other was written individually and included peer review.

Once it was known that the course would be conducted remotely, the final project was altered to focus on a measurement attainable at the students' homes that would still allow them to experience the entire experimentation process. The course materials instruct students to "measure something about the physical world around you, using only equipment and materials that you have available." The objective of the project was for the students to learn about the entire experimental process, from proposing a project to implementing it to presenting the results, including revisions along the way. In the course materials, the instructor emphasized to the students that the process was more important than the final result.

The students were given time early on in the term to create a plan for their projects, so they would have a well-thought-out strategy and the requisite equipment in order to perform experiments at home. The brainstorming process began by the students posting ideas to an online discussion forum, and the students, as well as faculty and staff in the department, then provided feedback about the scope of the projects. The students then wrote detailed project proposals, which included a literature review, a clearly defined project scope, and a plan for acquiring the necessary

equipment. This allowed them to be fully prepared at the start of the four weeks dedicated entirely to the projects.

The students participated in their final projects in a wide variety of ways due to the remote nature of the course. The students used a mix of materials found at their homes and materials shipped to them. Some of the groups chose to have similar parts shipped to all three students, so each could build their own apparatus. Other groups chose to divide up the work such that one or two students operated the equipment while the others conducted data analysis or simulations. The documentation of the experimental process also differed by group with some groups delegating one of the students to write the majority of the entries in their collective lab notebook, while other groups had all students share that responsibility. At the end of the final project, the students communicated their results in two formats: a lab report and a blog post targeting an audience of their peers. Both of these were written collaboratively by all group members, and the blog post was a new addition that year to replace a poster presentation from previous in-person versions of the course.

b. Course 2

Course 2 is an advanced lab course often taken by senior physics majors as a culminating lab course that provides the opportunity for more student agency than in the lower-division labs. This course is taught at a large, public, master's degree granting, Hispanic-serving institution in the United States. From the syllabus, the overall goal of the course is "to introduce physics students to advanced instrumentation, quantitative analysis, and realistic forms of communication used in physics and other scientific disciplines." In the winter term of 2020, there were four students enrolled in this course. Almost the entire course had occurred before the transition to remote learning due to the COVID-19 pandemic. Thus, all of the lab work was performed in person without any social distancing protocols.

The first half of the ten-week-long course consisted of structured experiments carried out by pairs of students. In the first two weeks, both pairs of students performed the same experiment where they used a current balance to measure the magnetic permeability constant. After that, each pair of students picked one out of four possible structured labs and worked on it for the following three weeks. One of the groups used scanning probe microscopes to image a diffraction grating and the other used a Geiger counter to characterize radioactivity. The students worked with the same lab partner throughout the entire course. Even though they worked in pairs, they wrote in their own lab notebooks and wrote individual lab reports, each of which had a peer review stage since writing was a focus of this course.

The second five weeks of the course were dedicated entirely to the final project, which was described in the course syllabus as "an experiment of [the students'] own

design.” The instructor wanted the final projects to be a chance for the students to learn how science is done, including data collection, data analysis, and the possibility to revise the experiment as needed. The students were required to submit a project proposal as a group early in the course, and the instructor provided feedback on the proposals to ensure the students were prepared to conduct their experiments.

The majority of the students’ work on the final projects occurred before the transition to remote learning, leading to the projects being carried out similarly to before the pandemic. The students in each group were allowed to be physically in the same location as each other while working on the project, and all of them worked in a lab. The students lost a couple of days in lab at the end of the project, but both of the projects were still able to obtain a measurement of at least one of their goal quantities. The students kept individual paper lab notebooks during the final projects (as well as the rest of the course), and each student wrote an entry for every day they were in lab. At the end of the projects, the students wrote individual final reports, peer reviewed final reports from the other group, and gave a joint presentation on their work.

c. Course 3

Course 3 is an advanced lab course typically for junior physics majors that provides the students a lab experience where they can synthesize their learning from prior courses. It is taught at a large, private, predominately white, doctoral-degree-granting research university in the United States. The course objectives are to use experimental systems relevant to contemporary physics, design projects, learn about proposal writing and evaluation, and present results. In the winter 2020 term, there were 21 students enrolled in this courses, divided into two sections. The course changed to being remote in the middle of the term due to the COVID-19 pandemic. After becoming remote, the learning goals remained mostly unchanged.

Prior to the final project, the students worked on two other lab experiments, which were structured, but still designed to allow students to build and test their own set-ups. The term was 15 weeks long, and each structured experiment lasted for approximately four weeks. One of these experiments involved vacuum systems and high voltages and the other involved optics and microfabrication. The students worked on these experiments in groups of two to three students with all of the students in a single section working on the same experiment at the same time. These structured experiments were completed before the transition to remote learning due to the COVID-19 pandemic, so the students were able to perform them in the lab. The experiments provided the students familiarity with some of the equipment they could use for their final projects.

The students in this course underwent a long process of creating and evaluating project proposals before beginning

construction of the project apparatus. In the course materials, the instructors described the goals of the final project to be that the students “demonstrate the scientific techniques and critical thinking that [they] have developed throughout the semester” and “learn how to identify, design, execute, and sell [their] ideas within a scientific community.” In order to enable the latter, all of the students individually identified a project topic and wrote a white paper about it approximately two months before the students began implementing their projects. Based on peer feedback, half of the white papers were chosen to be “funded” and each of those students paired up with a student whose white paper was not chosen. The pair of students then wrote a project proposal with the goal of fleshing out the plan for the experiment so the students would be prepared to work on it during the final four weeks of the term.

The transition to remote learning occurred during the beginning of the course time dedicated solely to the final projects, and each project was affected differently. Some of the students stayed in the location of the institution and were able to go in to lab one student at a time, while others had traveled farther away and were not able to use the institution’s lab equipment. Many of the students were forced to switch from their proposed projects to something that could be done remotely. Some of the students changed to doing a project they could access online, others brought equipment from campus to their homes, and yet others ended up doing a literature review. Many of the students, particularly ones who switched to remote projects, did not continue writing in their individual lab notebooks after the change. The instructor commented that many of the projects did not make as much progress towards their goal as anticipated due to impacts of the pandemic. The students gave virtual presentations about their projects to their classmates and instructors at the end of the term.

2. Data sources

There were two types of lab notebooks used in these courses: traditional paper lab notebooks were kept by each student in courses 2 and 3 and group electronic lab notebooks were kept by students in course 1. The paper lab notebooks contained entries for every day the students were in lab, although some students in course 3 stopped writing in their lab notebooks during the final project, presumably because of the disruption of the course due to the COVID-19 pandemic. Instead of daily entries, most of the electronic notebooks in course 1 had entries that summarized specific aspects of the projects. In some groups, one student submitted all of the entries to the lab notebook, while in other groups, all three students contributed approximately equally.

Each of the courses required a summative presentation of their project results in at least one format. Courses 1 and 2 both had written final reports. In course 2, these reports were done individually, so we had two lab reports for each

project. Course 3 required students to give a group oral presentation, so we analyzed the students' slides and a transcript of the video-recorded final presentations. Because the final presentations do not contain identical kinds of information as the final reports, we also use the final project proposals from course 3 for the projects that did not change due to COVID-19. We consider a change to have occurred when the students switched to an entirely new project topic or changed the modality (e.g., from building their own experiment in lab to remotely operating a publicly available apparatus), as evidenced by the differences between the proposal and the final presentation.

Student reflections are another important source of data to understand student ideas around modeling and engagement in modeling practices. The reflection questions were asked regularly throughout the courses and include questions about student experiences with their final project, as well as other aspects of the course. Example questions include [19,54]

- Describe a problem you experienced this week while working in the lab.
- What strategies did you and your group use to troubleshoot and solve the problem you encountered?
- What aspect of your contributions to the final project demonstrates your strengths and talents and why?

For our analysis, we only coded reflection questions that were given during, and pertained to, the final project.

The implementation of the student reflections varied by course. In course 1, reflection assignments were given as homework every couple of weeks throughout the term, each consisting of four to five questions with most students writing a paragraph in response to each question. We coded the final two reflections, since the earlier ones did not pertain to the final project. Sixteen out of the 24 students in the course opted-in to allow us to use their reflections in this analysis. In Courses 2 and 3, the student reflections consisted of weekly Qualtrics surveys. Each survey contained two to three open-response questions, and the student response lengths ranged from a short phrase to a few sentences. The students were provided the opportunity to opt in to the research study, and all but one student did. However, we are missing additional student reflections from these courses because not all students answered all of the questions every week. Both the response rate and the response length varied by week, question, and student.

We gathered additional information about students' projects through student interviews. We recruited students through emails to the entire class, explaining that the interviews were a chance for the students to reflect on their experiences in the course and to help improve future lab classes at their institution and nationally. The interviews were not connected to the students' course grades,

participation was voluntary, and the students were compensated for their time. We conducted interviews via Zoom of 10 students at the end of their courses. Three of the students interviewed were in course 1, two were in course 2, and five were in course 3. We asked each student several demographic questions, and found that two of the students were sophomores, five were juniors, and three were seniors. All of them were either physics or applied physics majors. We additionally asked students if they were willing to report their race, ethnicity, and gender, and all of them were willing to do so without prompted categories. Seven of the students self-identified as male, two self-identified as female, and one self-identified as transgender. Nine of the students self-identified as white and one self-identified as Hispanic/Mexican-American.

The interviews were semistructured with questions about the final projects as well as other aspects of the courses. The interviews ranged from 39–59 min, although only a portion of them were about the final projects. We coded the entire interviews since students discussed their final projects both when explicitly asked about them and when asked more general questions about the course. Relevant sample interview questions include the following:

- What aspects of experimental physics do you feel like you experienced in this class?
- Briefly describe your final project: what your project was about, what the goal was, how it went.
- What was the biggest challenge you encountered in doing the project?
- Do you feel like this diagram [of the EMF] reflects what you did during your final project?

The interview protocols were altered for the three courses (e.g., course 1 was asked more about the remote format), and each student was asked follow-up questions for clarification.

3. Ethics of research during pandemic

Once the pandemic started, we considered the ethics of continuing the study, concluded we could follow through with the study without detriment to the students, and adjusted our research plan accordingly. All of our data sources aside from the student interviews were already part of the course designs. The student interviews were optional, and we conducted them remotely and were flexible about the timing. We therefore did not add an additional burden to the students during this difficult time.

APPENDIX B: FEATURE CODES

Table V shows a complete list of the emergent feature codes with their definitions.

TABLE V. Project feature codes and their definitions.

Feature code	Definition
Build complex apparatus (A1)	The project provides the students the opportunity to build at least part of a complex apparatus. A complex apparatus is one that contains many components and requires careful alignment of those components, such that an expert would need to spend time aligning the setup even after the pieces were approximately in place. For example, this could involve a setup with optical or acoustic radiation.
Build intermediate-complexity apparatus (A2)	The project provides the students the opportunity to build at least part of an intermediate-complexity apparatus. An intermediate-complexity apparatus has approximately 5–10 parts, and at least some part of it will require some effort to assemble. However, it doesn't require careful alignment, so once it is fully put together it should work easily. This includes both electrical circuits and mechanical systems.
Build simple apparatus (A3)	The project provides the students the opportunity to build at least part of a simple apparatus. A simple apparatus involves only a few parts, such as a container with one material inside or a circuit consisting of 2-3 components. It should be able to be assembled by an expert in a few minutes with a high probability of working immediately.
No building (A4)	The project does not provide the students the opportunity to build any part of an apparatus. This could be because the students remotely operate a publicly available apparatus.
Use only lab equipment (B1)	All of the equipment required for the project is commonly found in a physics lab. Examples include optical elements (lenses, mirrors, etc.) and electronic parts (resistors, op amps, amplifiers, multimeters, etc.).
Use some equipment from home (B2)	Some of the equipment required for the project is more commonly found in a standard home than in a typical physics lab. Examples include plastic containers and sensors on a phone. The project may additionally require some equipment more commonly found in a lab.
Conceive and plan at least part of apparatus (C1)	The project provides the students the opportunity to come up with ideas for at least part of the apparatus and/or procedure themselves.
Don't conceive and plan apparatus (C2)	The project does not provide the students the opportunity to devise the plan for the majority of the apparatus; the students could just follow instructions when building it. This also includes when the students remotely control a publicly available apparatus without many options for setting it up.
Goal: Characterize apparatus (D1)	The primary goal of the project is to better understand a single apparatus (or several apparatus) including what properties it has and whether it would be better or worse than other options for a specific application.
Goal: Question with known solution (D2)	The project is confirmatory, where the primary goal is to measure some known quantity, either a single number or a relationship between two parameters with a known trend.
Goal: Question with specific instance unknown (D3)	The primary goal of the project is to measure some property that is known for other similar apparatus or other materials but students believe is not known for the specific case they are investigating.
Established model exists (E1)	It is possible for the goal of the project to be achieved and verified using a single established model (whether it is an equation or a dataset). This model can be easily looked up, and the students only need to do at most minor adaptations to make it align with their experiment (e.g., adding vectors).
Requires model construction (E2)	The project requires the students to do the majority of the model construction on their own in order to build the apparatus or have a theoretical prediction with which to compare their experimental results. There is no single place they can look up a model in the literature.

APPENDIX C: MODELING CODES

This section provides the definitions for all of the modeling codes used in our analysis. The codes describing the main modeling tasks of the EMF are defined in Table VI, with the emergent subcodes describing engagement with models shown in Table VII and the emergent subcodes describing revisions shown in Table VIII. Any project assigned one of the subcodes was also assigned the code for the corresponding main task.

TABLE VI. Codes for the main tasks of the Experimental Modeling Framework and their definitions. These are labeled as “main modeling codes” in Fig. 2.

Modeling code	Definition
Make measurement	The students take a raw measurement.
Engage with model of measurement system	The students engage with a model of the measurement system. The measurement system is all parts of the apparatus needed not to define the goal question but only to answer it. Engagement can include constructing a model or discussing limitations of the measurement device or model, but it needs to be more than the students just stating what device they use.
Engage with model of physical system	The students engage in any capacity with a model of the physical system. The physical system is everything needed to create whatever system the goal question is about (but not necessarily to answer it). This includes the students finding a model in the literature (without any mention of how it needs to be adjusted for their specific instantiation) or putting together different pieces of the model themselves. This also includes general comments about modeling their system.
Analyze data	The students convert the data from its raw form to a form that can be compared with a theoretical prediction. This includes everything from cleaning up the data (interpolating, deciding which data to use, etc.) to plugging into equations from either the measurement system model or the physical system model to extract the desired quantity.
Make comparison	The students compare experimental data with a prediction from a theoretical model. This comparison could be of anything ranging from a quantitative measurement of the goal quantity to a qualitative troubleshooting check along the way.
Propose causes	The students propose a cause for a discrepancy between an experimental result and a theoretical prediction. This discrepancy could come from a quantitative measurement or a qualitative troubleshooting check.
Revise measurement apparatus	The students change some part of the measurement system apparatus after they have already implemented it. The measurement system apparatus is all parts of the apparatus used not to define the project goal question but only to answer it.
Revise measurement model	The students change some part of the model of the measurement system after they have already applied it. The measurement model is the model of all parts of the experiment used not to define the project goal question but only to answer it. This includes revising the model of the measurement apparatus, revising the method of data analysis, and revising the measurement procedure.
Revise physical apparatus	The students change some part of the physical system apparatus after they have already implemented it. The physical system apparatus is all parts of the apparatus needed to create whatever system the project goal question is about (but not necessarily to answer it).
Revise physical model	The students change some part of the physical system model after they have already applied it. The physical system model is all parts of the model related to the parts of the apparatus needed to create whatever system the project goal question is about (but not necessarily to answer it).

TABLE VII. Emergent subcodes of *engage with models* codes and their definitions. The first two codes in this table are subcodes of *engage with model of measurement system* and the last four are subcodes of *engage with model of physical system*.

Modeling code	Definition
Construct model of measurement system	The students construct a model of the measurement system.
Discuss limits of measurement system	The students mention limitations or assumptions of the measurement system (including the measurement model, the measurement apparatus, and data analysis), whether or not they explicitly discuss how these will affect their project.
Construct model of physical system	The students construct, or attempt to construct, a model of the physical system.
Discuss limits of physical system	The students mention limitations or assumptions of the physical system, whether or not they explicitly discuss how it will affect their project. This includes observations of the way their physical system is imperfect compared to the ideal model, limits of the physical apparatus itself, and assumptions in the model.

(Table continued)

TABLE VII. (*Continued*)

Modeling code	Definition
Simulate or calculate parts of physical system model	The students take an already existing model and apply it to their specific setup. This involves some thought from the students and may consist of deciding which parts of an equation to use, putting several equations together, figuring out what range of values is acceptable for their experiment, adjusting for the geometry of their setup, or playing around with a simulation someone else created.
Write down physical system model	The students write down the accepted model from the literature without relating it to their own project.

TABLE VIII. Emergent subcodes of the four revisions codes along with their definitions.

Modeling code	Definition
Revise measurement apparatus—major	The students revise the measurement system apparatus in a major way, by adding, removing, or switching out a part of the apparatus.
Revise measurement apparatus—minor	The students revise the measurement system apparatus in a minor way, by slightly altering the already existing apparatus. Examples include playing with the apparatus' settings or physically adjusting its orientation or location.
Revise measurement model—major	The students revise the measurement system model in a major way, by adding, removing, or switching out parts of the model. This includes entirely changing the measurement procedure.
Revise measurement model—minor	The students revise the measurement system model in a minor way, by slightly altering the already existing model. Examples include fixing a small mathematical mistake or altering the measurement procedure by measuring under a slightly different condition.
Revise measurement model—data analysis	The students revise how they analyze their data or calculate uncertainty. This includes revising any part of their methods after the raw measurement has been recorded and up to when the students have the result needed to compare with a theoretical prediction.
Revise physical apparatus—major	The students revise the physical system apparatus by adding, removing, or switching out a part of the apparatus such that they need to consider how their system would work after the change.
Revise physical apparatus—minor	The students revise the physical system apparatus by adjusting the already-existing parts. This includes basic troubleshooting such as re-aligning optics, playing with parameters and settings, and re-connecting parts correctly.
Revise physical apparatus—unknown	The students revise the physical system apparatus, but there is not enough context to know how small or big this revision is.
Revise physical model—major	The students revise the physical system model by switching to an entirely new model.
Revise physical model—minor	The students revise the physical system model by making a small change, either correcting a mistake they had previously made or making minor tweaks to the model.

APPENDIX D: MODEL ENGAGEMENT AND REVISIONS BY PROJECT

This section contains the data (Fig. 5) describing the prevalence of the *engage with models* and *revisions* subcodes across the projects, the results of which were summarized in Sec. VA. Note that it is possible for the projects that did not build anything to make revisions to the apparatus because we consider the programming code used to run the experiments a part of the apparatus.

(a)	Acoustic Levitation	Blackbody Radiators	Chain Fountain	Earth's Magnetic Field	IBM Quantum Computer	Model Rockets	Optical Pumping	Paschen Curve	Phonons	Photon Detector	Refractive Index	Sonoluminescence	Tensile Strength Yarn	Van der Pol Oscillator	N
Construct model of measurement system	No	No	No	No	No	Yes	No	No	No	No	No	No	Yes	No	2
Discuss limits of measurement system	No	No	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	No	Yes	No	7
Construct model of physical system	No	No	Yes	No	Yes	Yes	No	No	No	No	No	No	No	Yes	4
Discuss limits of physical system	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	9
Simulate or calculate parts of physical system model	Yes	Yes	Yes	Yes	No	Yes	No	No	No	No	No	Yes	No	No	6
Write down physical system model	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	14

(b)	Acoustic Levitation	Blackbody Radiators	Chain Fountain	Earth's Magnetic Field	IBM Quantum Computer	Model Rockets	Optical Pumping	Paschen Curve	Phonons	Photon Detector	Refractive Index	Sonoluminescence	Tensile Strength Yarn	Van der Pol Oscillator	N
Revise measurement apparatus - major	No	Yes	No	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	9
Revise measurement apparatus - minor	No	No	Yes	Yes	No	No	Yes	No	No	No	No	No	No	No	3
Revise measurement model - major	No	No	No	No	No	Yes	No	No	No	No	No	No	No	No	1
Revise measurement model - minor	No	No	No	Yes	No	No	No	No	No	No	No	No	Yes	No	2
Revise measurement model - data analysis	No	No	Yes	No	No	Yes	No	No	No	No	Yes	No	Yes	Yes	5
Revise physical apparatus - major	Yes	No	Yes	No	No	No	No	No	Yes	Yes	No	No	Yes	Yes	6
Revise physical apparatus - minor	Yes	No	Yes	No	No	Yes	Yes	No	Yes	Yes	No	Yes	No	No	7
Revise physical apparatus - unknown	No	No	No	No	Yes	No	No	No	No	No	No	No	No	No	1
Revise physical model - major	No	No	No	No	No	Yes	No	No	No	No	No	No	No	No	1
Revise physical model - minor	No	No	Yes	No	No	No	No	No	No	No	Yes	No	No	No	2

FIG. 5. Presence (“yes,” blue) or absence (“no,” white) of emergent subcodes for the main modeling codes (a) *engage with models* and (b) *revisions* in each of the 14 projects. The number of projects that show evidence of each subcode are listed in the far-right column.

APPENDIX E: ASSOCIATIONS BETWEEN REVISIONS AND PROJECT FEATURES

Figure 6 shows the co-occurrence of the project features and revisions subcodes. Instead of grouping all the revisions subcodes together, we have divided them into four separate plots, showing the subcodes related to (a) apparatus revisions, (b) model revisions, (c) physical system revisions, and (d) measurement system revisions. Each of the subcodes appears in two of these categories. This division allows us to investigate how the features of the projects may have led to the students revising different aspects of their project. During a several-week-long project, students will not have sufficient time to enact all possible types of revisions, and instructors may want their students to focus on only one of these categories.

Model revisions exhibit a wide variation by project feature, as is evidenced by many of the columns in Fig. 6(b) being entirely white, while many other columns are at least partially shaded. The features of projects in which students most commonly enacted model revisions are the same as the ones in which students performed model

construction (see Table IV). This similarity is expected because major model revisions are often additionally coded as model construction since a major revision of a model is often the same as constructing a new model. Model construction can occur at any time, whereas model revisions can only occur after there has been a first attempt previously implemented.

We also find some differences across project features for measurement system revisions [see Fig. 6(d)], which are dominated by the differences in revisions of the measurement models. Many of the projects are coded as doing *revise measurement apparatus—major*, so there is less of a feature dependence for that kind of revision. However, the other measurement system revisions show up in projects with only a subset of the features. Most of the project features most associated with measurement system revisions overall are similar to those associated with model construction and model revisions (see Table IV), although the feature *requires model construction* is less associated. In fact, projects that require model construction and projects that use already established models have similar column averages for measurement system revisions.

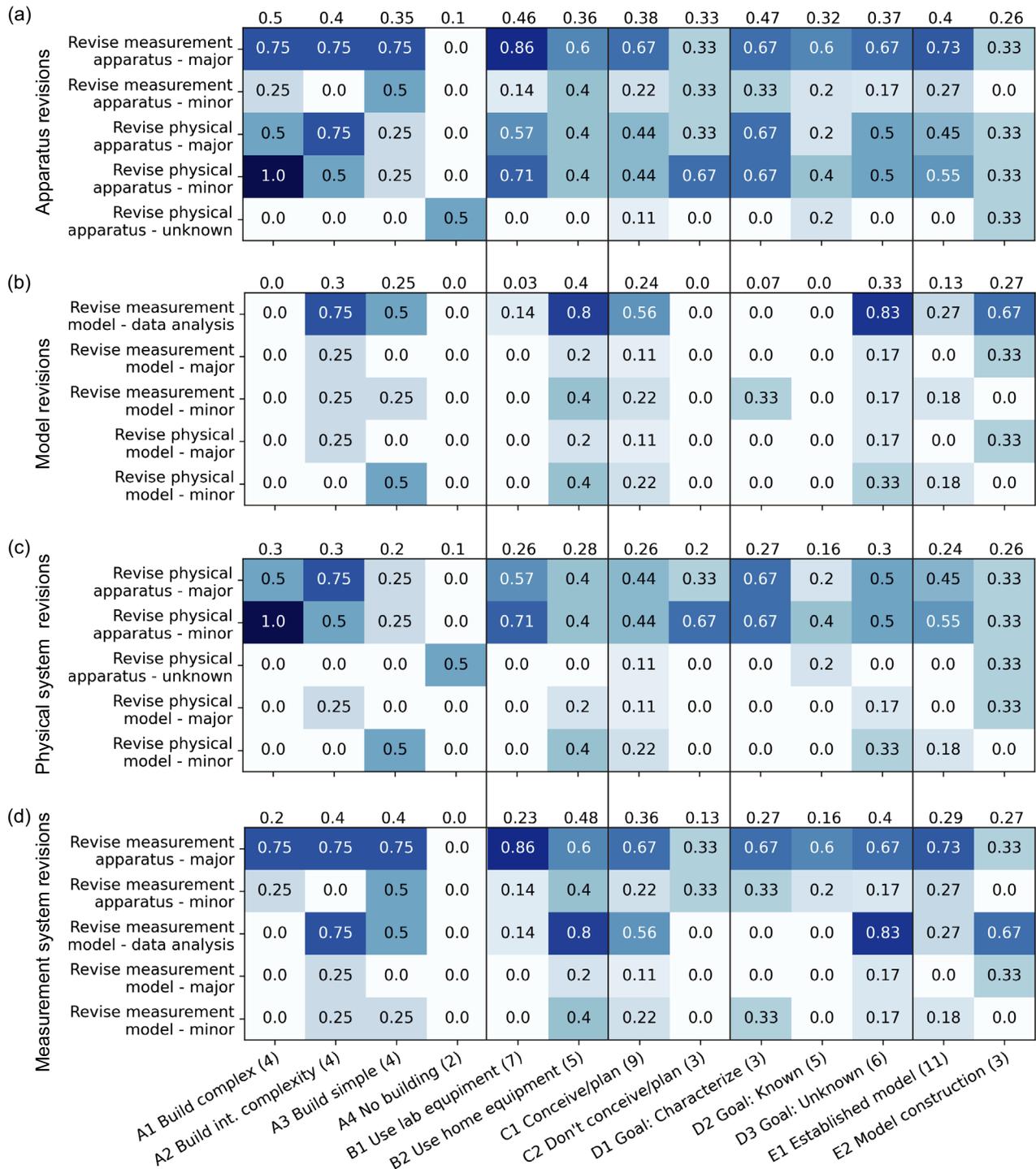


FIG. 6. Fraction of projects with a specified feature (columns, organized by category) assigned the *enact revisions* subcodes (rows) for (a) apparatus revisions, (b) model revisions, (c) physical system revisions, and (d) measurement system revisions. The numbers in parentheses following the abbreviated feature names are the number of projects with each feature, and the numbers on top are averages over each column. For each of the subplots, the features with the largest column averages can be interpreted as being most associated with that category of revisions.

The similarity between the features associated with model and measurement system revisions may be partially due to this dataset where some of the measurement system model construction and measurement system revisions

arose from the students being at home without common measurement devices.

One feature that stands out in all categories of revisions is *no building* because it co-occurs with only one of the

revision subcodes. This is unsurprising since the students that did not build an apparatus were not able to change the apparatus itself and were only able to revise either the commands sent to the apparatus or the model of it. However, it is important to note that both projects with this feature had

to change last minute, so the lack of revisions could also come about from having less time to work on the project. Prior work has hypothesized that students may not proceed to the revision stage even if they identified a need for it when the students are rushing to finish the lab [47].

-
- [1] J. Kozminski, H. Lewandowski, N. Beverly, S. Lindaas, D. Dearthoff, A. Reagan, R. Dietz, R. Tagg, M. Eblen-Zayas, J. Williams *et al.*, *AAPT recommendations for the undergraduate physics laboratory curriculum* (American Association of Physics Teachers, College Park, MD, 2014).
- [2] C. V. Schwarz, B. J. Reiser, E. A. Davis, L. Kenyon, A. Achér, D. Fortus, Y. Shwartz, B. Hug, and J. Krajcik, Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners, *J. Res. Sci. Teach.* **46**, 632 (2009).
- [3] E. Etkina, A. Warren, and M. Gentile, The role of models in physics instruction, *Phys. Teach.* **44**, 34 (2006).
- [4] B. M. Zwickl, D. Hu, N. Finkelstein, and H. J. Lewandowski, Model-based reasoning in the physics laboratory: Framework and initial results, *Phys. Rev. ST Phys. Educ. Res.* **11**, 020113 (2015).
- [5] C. Wieman, Comparative cognitive task analyses of experimental science and instructional laboratory courses, *Phys. Teach.* **53**, 349 (2015).
- [6] D. Hestenes, Toward a modeling theory of physics instruction, *Am. J. Phys.* **55**, 440 (1987).
- [7] Joint Task Force on Undergraduate Physics Programs, Phys21: Preparing physics student for 21st-century careers (2016).
- [8] D. R. Dounas-Frazer and H. J. Lewandowski, Nothing works the first time: An expert experimental physics epistemology, in *PERC Proceedings 2016* (American Association of Physics Teachers, Sacramento, 2016), pp. 100–103, [10.1119/perc.2016.pr.020](https://doi.org/10.1119/perc.2016.pr.020).
- [9] D. R. Dounas-Frazer, J. T. Stanley, and H. Lewandowski, Instructor perspectives on iteration during upper-division optics lab activities, in *Proceedings of the Physics Education Research Conference, Cincinnati, OH, 2017* (AIP, New York, 2017), pp. 116–119, [10.1119/perc.2017.pr.024](https://doi.org/10.1119/perc.2017.pr.024).
- [10] N. M. Juma, E. Gire, K. Corwin, B. Washburn, and N. S. Rebello, Students' and instructor's impressions of ill-structured capstone projects in an advanced electronics lab, *AIP Conf. Proc.* **1289**, 181 (2010).
- [11] N. G. Holmes and C. E. Wieman, Examining and contrasting the cognitive activities engaged in undergraduate research experiences and lab courses, *Phys. Rev. Phys. Educ. Res.* **12**, 020103 (2016).
- [12] J. R. Hoehn and H. J. Lewandowski, Incorporating writing in advanced lab projects: A multiple case-study analysis, *Phys. Rev. Phys. Educ. Res.* **16**, 020161 (2020).
- [13] J. R. Hoehn and H. J. Lewandowski, Framework of goals for writing in physics lab classes, *Phys. Rev. Phys. Educ. Res.* **16**, 010125 (2020).
- [14] J. T. Stanley, D. R. Dounas-Frazer, L. Kieपुरa, and H. J. Lewandowski, Investigating student ownership of projects in an upper-division physics lab course, in *Proceedings of the Physics Education Research Conference, Sacramento, CA* (AIP, New York, 2016), pp. 336–339, [10.1119/perc.2016.pr.079](https://doi.org/10.1119/perc.2016.pr.079).
- [15] D. R. Dounas-Frazer, J. T. Stanley, and H. J. Lewandowski, Student ownership of projects in an upper-division optics laboratory course: A multiple case study of successful experiences, *Phys. Rev. Phys. Educ. Res.* **13**, 020136 (2017).
- [16] D. R. Dounas-Frazer, L. Ríos, and H. J. Lewandowski, Preliminary model for student ownership of projects, in *Proceedings of the 2019 Physics Education Research Conference, Provo, UT* (AIP, New York, 2019), [10.1119/perc.2019.pr.Dounas-Frazer](https://doi.org/10.1119/perc.2019.pr.Dounas-Frazer).
- [17] I. C. Lassen, A. Arielle-Evans, L. Ríos, H. Lewandowski, and D. Dounas-Frazer, Student ownership and understanding of multiweek final projects, in *Proceedings of PER Conf. 2021, virtual conference*, [10.1119/perc.2021.pr.Lassen](https://doi.org/10.1119/perc.2021.pr.Lassen).
- [18] B. R. Wilcox and H. J. Lewandowski, Open-ended versus guided laboratory activities: Impact on students' beliefs about experimental physics, *Phys. Rev. Phys. Educ. Res.* **12**, 020132 (2016).
- [19] M. Eblen-Zayas, The impact of metacognitive activities on student attitudes towards experimental physics, in *Proceedings of the Physics Education Research Conference 2016, Sacramento, CA* (AIP, New York, 2016), pp. 104–107, [10.1119/perc.2016.pr.021](https://doi.org/10.1119/perc.2016.pr.021).
- [20] D. R. Dounas-Frazer and H. Lewandowski, The Modelling Framework for Experimental Physics: Description, development, and applications, *Eur. J. Phys.* **39**, 064005 (2018).
- [21] 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).
- [22] N. G. Holmes and C. E. Wieman, Introductory physics labs: We can do better, *Phys. Today* **71**, No. 1, 38 (2018).
- [23] N. G. Holmes and H. J. Lewandowski, Investigating the landscape of physics laboratory instruction across North America, *Phys. Rev. Phys. Educ. Res.* **16**, 020162 (2020).
- [24] E. M. Smith and N. Holmes, Best practice for instructional labs, *Nat. Phys.* **17**, 662 (2021).

- [25] L. B. Buck, S. L. Bretz, and M. H. Towns, Characterizing the level of inquiry in the undergraduate laboratory, *J. Coll. Sci. Teach.* **38**, 52 (2008), <https://cursa.ihmc.us/rid=1Q3XD4DZ3-1WDCR3Y-F1VK/Levels%20of%20Inquiry%20Undergrad.pdf>.
- [26] N. Holmes, C. E. Wieman, and D. Bonn, Teaching critical thinking, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11199 (2015).
- [27] A. Price, C. Kim, E. Burkholder, A. Fritz, and C. Wieman, A universal structure of science and engineering problem-solving, *CBE—Life Sci. Edu.* **20**, ar43 (2021).
- [28] D. Hu, B. M. Zwickl, B. R. Wilcox, and H. J. Lewandowski, Qualitative investigation of students' views about experimental physics, *Phys. Rev. Phys. Educ. Res.* **13**, 020134 (2017).
- [29] P. W. Irving and E. C. Sayre, Conditions for building a community of practice in an advanced physics laboratory, *Phys. Rev. ST Phys. Educ. Res.* **10**, 010109 (2014).
- [30] N. G. Holmes, B. Keep, and C. E. Wieman, Developing scientific decision making by structuring and supporting student agency, *Phys. Rev. Phys. Educ. Res.* **16**, 010109 (2020).
- [31] F. Pols, A physics lab course in times of COVID-19, *Elec. J. Res. Sci. Math. Educ.* **24**, 172 (2020), <https://eric.ed.gov/?id=EJ1261593>.
- [32] B. Cai, L. A. Mainhood, R. Groome, C. Laverty, and A. McLean, Student behavior in undergraduate physics laboratories: Designing experiments, *Phys. Rev. Phys. Educ. Res.* **17**, 020109 (2021).
- [33] J. R. Hoehn, M. F. J. Fox, A. Werth, V. Borish, and H. J. Lewandowski, Remote advanced lab course: A case study analysis of open-ended projects, *Phys. Rev. Phys. Educ. Res.* **17**, 020111 (2021).
- [34] Z. Y. Kalender, E. Stump, K. Hubenig, and N. G. Holmes, Restructuring physics labs to cultivate sense of student agency, *Phys. Rev. Phys. Educ. Res.* **17**, 020128 (2021).
- [35] I. Halloun, Schematic modeling for meaningful learning of physics, *J. Res. Sci. Teach.* **33**, 1019 (1996).
- [36] E. Etkina and A. Van Heuvelen, *Investigative science learning environment—A science process approach to learning physics* (American Association of Physics Teachers, College Park, MD, 2007), pp. 1–48.
- [37] E. Brewster, Modeling theory applied: Modeling Instruction in introductory physics, *Am. J. Phys.* **76**, 1155 (2008).
- [38] P. D. O. Bergeron, A. Sirnoorkar, and J. Laverty, Towards an analytic framework for characterizing student use of models, in *Proceedings of PER Conf. 2021, virtual conference*, 10.1119/perc.2021.pr.Bergeron.
- [39] D. McPadden and E. Brewster, Impact of the second semester University Modeling Instruction course on students' representation choices, *Phys. Rev. Phys. Educ. Res.* **13**, 020129 (2017).
- [40] C. Walsh, K. N. Quinn, C. Wieman, and N. G. Holmes, Quantifying critical thinking: Development and validation of the physics lab inventory of critical thinking, *Phys. Rev. Phys. Educ. Res.* **15**, 010135 (2019).
- [41] L. Bao, K. Koenig, Y. Xiao, J. Fritchman, S. Zhou, and C. Chen, Theoretical model and quantitative assessment of scientific thinking and reasoning, *Phys. Rev. Phys. Educ. Res.* **18**, 010115 (2022).
- [42] D. R. Dounas-Frazer, L. Ríos, B. Pollard, J. T. Stanley, and H. J. Lewandowski, Characterizing lab instructors' self-reported learning goals to inform development of an experimental modeling skills assessment, *Phys. Rev. Phys. Educ. Res.* **14**, 020118 (2018).
- [43] L. Ríos, B. Pollard, D. R. Dounas-Frazer, and H. J. Lewandowski, Using think-aloud interviews to characterize model-based reasoning in electronics for a laboratory course assessment, *Phys. Rev. Phys. Educ. Res.* **15**, 010140 (2019).
- [44] B. Pollard, M. F. Fox, L. Ríos, and H. Lewandowski, Creating a coupled multiple response assessment for modeling in lab courses, in *Proceedings of PER Conf. 2020, virtual conference*, 10.1119/perc.2020.pr.Pollard.
- [45] M. F. Fox, B. Pollard, L. Ríos, and H. Lewandowski, Capturing modeling pathways using the Modeling Assessment for Physics Laboratory Experiments, in *Proceedings of PER Conf. 2020, virtual conference*, 10.1119/perc.2020.pr.Fox.
- [46] D. R. Dounas-Frazer, K. L. Van De Bogart, M. R. Stetzer, and H. J. Lewandowski, Investigating the role of model-based reasoning while troubleshooting an electric circuit, *Phys. Rev. Phys. Educ. Res.* **12**, 010137 (2016).
- [47] J. T. Stanley, W. Su, and H. J. Lewandowski, Using lab notebooks to examine students' engagement in modeling in an upper-division electronics lab course, *Phys. Rev. Phys. Educ. Res.* **13**, 020127 (2017).
- [48] J. R. Hoehn and H. J. Lewandowski, Investigating undergraduate students' views about the process of experimental physics (to be published).
- [49] 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).
- [50] M. J. Farrar, Arduino™ microprocessors as a platform for modeling framework labs, *Am. J. Phys.* **89**, 596 (2021).
- [51] M. Eisenhart, *Generalization from qualitative inquiry, Generalizing from Educational Research: Beyond Qualitative and Quantitative Polarization* (Routledge, London, 2009), pp. 51–66.
- [52] A. D. Robertson, S. B. McKagan, and R. E. Scherr, Selection, generalization, and theories of cause in physics education research: Connecting paradigms and practices, *Getting Started in PER*, edited by C. Henderson and K. A. Harper (American Association of Physics Teachers, College Park, MD, 2018), Vol. 2, <http://www.per-central.org/items/detail.cfm?ID=14727>.
- [53] M. Eblen-Zayas and R. C. Terrien, Lessons learned from five years of student self-directed experimental projects in the advanced lab, in *Proceedings of the 2018 Conference on Laboratory Instruction Beyond the First Year of College, Baltimore, MD (2018)*, 10.1119/bfy.2018.pr.003.
- [54] D. R. Dounas-Frazer and D. L. Reinholz, Attending to lifelong learning skills through guided reflection in a physics class, *Am. J. Phys.* **83**, 881 (2015).