Using think-aloud interviews to characterize model-based reasoning in electronics for a laboratory course assessment

Laura Ríos,[*] Benjamin Pollard, Dimitri R. Dounas-Frazer,[†] and H. J. Lewandowski

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

(Received 7 January 2019; published 12 June 2019)

Models of physical systems are used to explain and predict experimental results and observations. The Modeling Framework for Experimental Physics describes the process by which physicists revise their models to account for the newly acquired observations, or change their apparatus to better represent their models when they encounter discrepancies between actual and expected behavior of a system. While modeling is a nationally recognized learning outcome for undergraduate physics lab courses, no assessments of students' model-based reasoning exist for upper-division labs. As part of a larger effort to create two assessments of students' modeling abilities, we used the Modeling Framework to develop and code think-aloud problem-solving activities centered on investigating an inverting amplifier circuit. This study is the second phase of a multiphase assessment instrument development process. Here, we focus on characterizing the range of modeling pathways students employ while interpreting the output signal of a circuit functioning far outside its recommended operation range. We end by discussing four outcomes of this work: (i) Students engaged in all modeling subtasks, and they spent the most time making measurements, making comparisons, and enacting revisions; (ii) each subtask occurred in close temporal proximity to all other subtasks; (iii) sometimes, students propose causes that do not follow from observed discrepancies; (iv) similarly, students often rely on their experiential knowledge and enact revisions that do not follow from articulated proposed causes.

DOI: 10.1103/PhysRevPhysEducRes.15.010140

I. INTRODUCTION

Recently, the National Research Council (NRC) called for increased attention to assessments of experimental physics practices, and to the assessment development process in particular [1]. One important experimental physics practice is modeling: the construction, testing, use, and revision of models of physical phenomena and apparatus. According to the recommendations released by the American Association of Physics Teachers (AAPT), modeling should be a focus of physics laboratory courses [2].

Currently, there is no published, validated instrument to assess modeling in physics laboratory courses. There are assessments that measure other concepts in labs, such as "critical thinking" [3], how students perceive experimental uncertainty [4], and students' views about experimental physics [5]. There are also several assessments for laboratory courses in science, technology, engineering, and math generally (see, e.g., Refs. [6,7] for examples from chemistry and biology). Prompted partly by the recent NRC call to action and the AAPT guidelines for labs, our approach to this work is to build on an assumption that labs are inherently valuable learning spaces, and that there is a need to invest in their improvement [8]. The creation of research-based assessments is a critical step in the process to realize this improvement.

To address the national calls by NRC and AAPT, our team is undertaking an extensive process to create two scalable assessments for model-based reasoning in physics lab courses that will be generalizable for use at various levels and institutions. There will be one assessment each for electronics and optics courses. The assessment development process has four phases: (i) determine domain-specific test objectives [9]; (ii) characterize how students navigate a lab-practicum style activity and justify their choices during think-aloud problem-solving (TAPS) interviews; (iii) create a free-response assessment with input from the TAPS interviews and expert physicists; and finally, (iv) create a closed-response format assessment

[*]laura.rios@colorado.edu
[†]Present address: Department of Physics and Astronomy, Western Washington University, Bellingham, Washington 98225, USA.

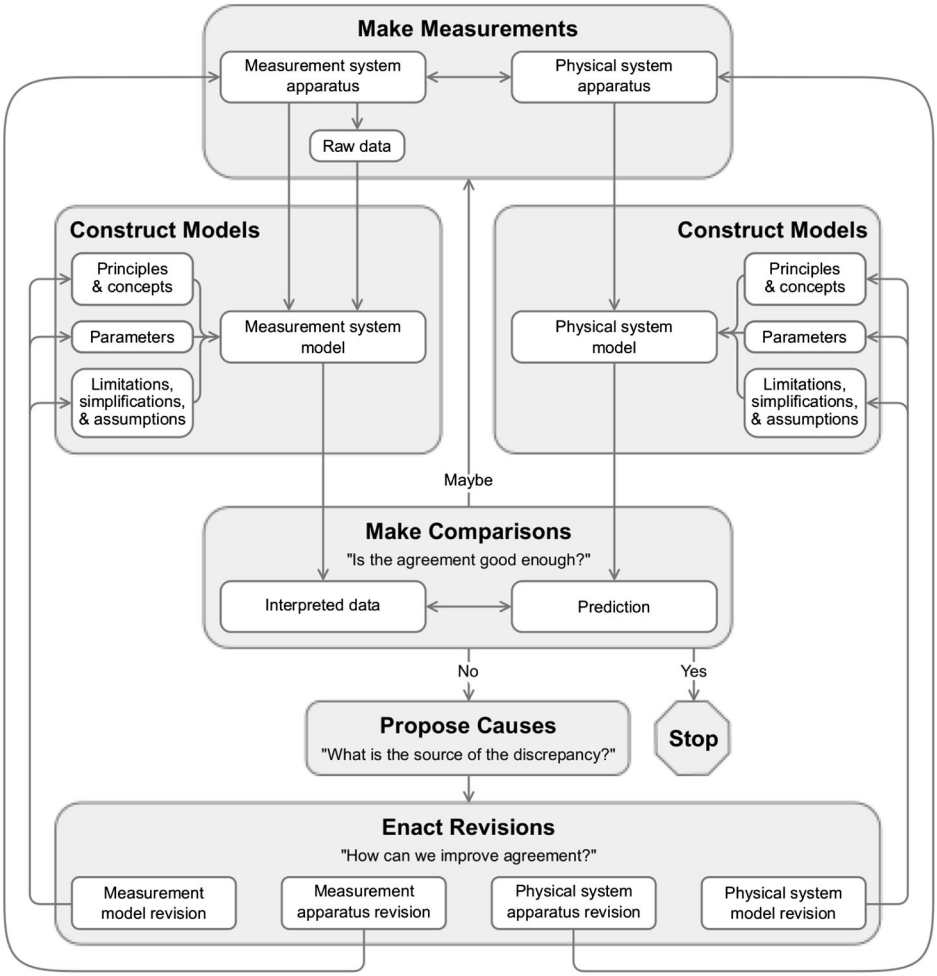

FIG. 1. Modeling Framework for Experimental Physics: Originally conceptualized by Zwickl *et al.* [12], this schematic of the Modeling Framework from Dounas-Frazer [9] shows how five interconnected tasks—construct models, make measurements, make comparisons, propose causes, and enact revisions—describe the ways scientists bring data and predictions from models into agreement. In the Modeling Framework, we distinguish between physical models or apparatus (e.g., models of the op-amp) and measurement models or apparatus (e.g., models of digital multimeters).

based on students responses from phase 3. The final instrument will have a similar deployment model as the Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS) assessment, resulting in a centralized online survey tool with automated administration [10].

We are using the Modeling Framework for Experimental Physics to guide all phases of the assessment development. The framework describes the process by which physicists bring measurements and predictions from models into agreement [11]. The framework is composed of several interconnected subtasks, depicted in gray boxes in Fig. 1: making measurements, constructing models, making comparisons between data and predictions to identify discrepancies, proposing causes for those discrepancies, and enacting revisions to resolve them.

In this paper, we describe the findings from phase 2 of the assessment development. In phase 2, we used outcomes from phase 1 to develop a hands-on electronics activity designed to elicit a large range of modeling behavior centered on the functional limits of operational amplifiers (op-amps). We conducted 10 TAPS interviews with upper-division physics and engineering physics undergraduate students at the University of Colorado Boulder (CU Boulder). After coding and analyzing the data, we determined several concrete implications for phases 3 and 4.

II. RESEARCH QUESTIONS

The research questions for phase 2 were developed looking ahead to the structure and content we required for phase 3. Specifically, we needed to understand how students navigate the modeling process to create relevant test items. We also incorporated instructor feedback from phase 1 to make sure our final instrument and associated findings aligned with instructor values.

With that in mind, our research goals were distilled to the following questions about students' approaches to modeling during the TAPS interviews:

RQ1. When students work on an op-amp circuit, how much time do they spend on each of the modeling subtasks?

RQ2. When students work on an op-amp circuit, which pairs of subtasks are temporally connected?

RQ3. When students work on an op-amp circuit, what types of apparatus and model revisions, comparisons, proposed causes, and measurements do the students make?

We want our final assessment to reflect what students spend time on, but we should not exclude less common subtasks, as they may still play a crucial role in the modeling process. Further, our final assessment instrument will need to have discriminatory power, i.e., differentiate between different levels of modeling ability.

Similarly, we also want to identify common, missing, or unexpected modeling pathways. Accordingly, RQ1 and RQ2 together seek to describe which subtasks students engage in most commonly, and to characterize typical pathways within the context of the Modeling Framework. RQ3 will provide important contextual information for specific scenarios and test items to be developed in phase 3, the free-response assessment. We refer to these three research questions throughout, and organize the results and discussion sections according to outcomes from each research question.

III. SELECT BACKGROUND RESEARCH

A. Research on modeling and measurements

Model-based reasoning has been researched extensively in physics education research, resulting in various definitions and conceptualizations of the modeling process and its role in physics (e.g., Refs. [13,14]). Here, we describe a few studies whose findings align with or complement our operationalization of modeling.

Some previous research in modeling focuses on the role of measurements. Koponen and others have discussed the link between predictions and measurements [15], and have gone further in describing *generative* modeling, wherein measurements are conceptualized as "investigative" [16]. In this context, investigative measurements can be used to probe concepts, models, and expectations throughout the modeling process, thereby generating knowledge.

Russ and Odden's work also focuses on measurements. They describe how students combine evidence-based reasoning and modeling to learn about physical phenomena, and work through initially confusing problems. One of their findings is that students use evidence to further elaborate on model components, and that students' models help guide the search for new evidence [17].

In our work, we combine Russ and Odden's findings on students' use of evidence with Koponen's description of investigative measurements to situate the role of measurements.

Another way of conceptualizing an investigative measurement is to understand its outcome. Along this vein, Vonk *et al.* define "applying a model" to express a quantitative relationship between variables as a part of "model making," and the necessary revision to the model after critical testing as "model breaking," both crucial scientific skills [18]. Students in that study were primed to consider either model making or model breaking while working on web-based lab activities. For our study, we expect both model-making and model-breaking behavior. An obvious difference is that our work will be conducted with students working on physical apparatus, not an online module, so we may expect more interaction with models of the apparatus itself.

Similarly, Allie and colleagues have grappled with students' understanding of the role of measurements, and how to compare two sets of measurements to one another [4,19–21]. In one particular study, Volkwyn and colleagues explored what physics majors in traditional introductory physics lab courses think about the nature of a scientific measurement [22]. One of their findings centered on how and why students came to understand if datasets were in agreement, an area referred to as data comparison. Using the physics measurement questionnaire (PMQ) to determine how students' thinking in this area shifted before and after lab instruction, they found that students' thinking shifted the least toward more favorable reasoning in data comparison when compared to how student reasoning shifted for data collection and data processing. In the Modeling Framework, the make comparisons subtask determines the stop criteria, so we are generally interested in ways that these two frameworks coalesce in the realm of data comparison.

B. Research on the Modeling Framework

Our own research into aligning lab courses with authentic physics experience and practice led to the development and use of the Modeling Framework for Experimental Physics (Fig. 1) in physics laboratory courses and clinical settings at the University of Colorado Boulder [11,12,23–26]. For a detailed summary of the development and applications of the Modeling Framework, see Ref. [27].

Particularly important for our study is previous work that has identified the propose causes subtask as particularly difficult. Zwickl and colleagues found that students' relatively limited theoretical domain knowledge (i.e., physics principles and concepts) hindered their ability to fully explore the limitations and assumptions of their models while working with a photodiode [11]. In a recent publication, Ríos *et al.* unpacked the varied ways students skip the propose causes subtask in the Modeling Framework in favor of other modeling subtasks [28]. That work aligns with our findings here, and we will show there is a need to understand the circumstances under

which a student does or does not propose a probable cause to motivate revisions.

Further, difficulty in proposing causes can create difficulty in other aspects of modeling, or interrupt the modeling process altogether. Another study on how students document modeling in their lab notebooks found that even when the modeling process is scaffolded by course materials, the students "generally did not provide [in their lab notebooks] actionable ways of implementing these proposed revisions" [24]. This indicates a difficulty in the enact revisions subtask stemming from an unarticulated proposed cause.

C. Troubleshooting in electronics

Previous research on students working on circuits tended to focus on conceptual understanding of electronics physics concepts (e.g., Ref [29]). Researchers have also observed how socially mediated metacognition plays a role in troubleshooting malfunctioning electronics circuits [30].

In another related study, students' model-based reasoning was observed while troubleshooting electric circuits [25]. Dounas-Frazer *et al.* describe the modeling process during troubleshooting as involving decisions about "which measurement to perform, in what order, and for what purpose" (p. 5). In their description of troubleshooting episodes, the initial formulation of the problem description is characterized partly by understanding the issues using formative measurements, testing the apparatus using diagnostic measurements to determine the source of a discrepancy, repairing the apparatus, and finally establishing that the overall function of the circuit is correct using evaluative measurements. Our findings here show similar ways of using different types of measurements.

We also draw on this work to inform our activity design. Dounas-Frazer *et al.* incorporated errors in circuit construction in their activity design, which led to several iterative cycles of figuring out how to fix the circuit. In addition to comparing their predictions to the observed output signal, the students also had to repair the circuit. To explore a different space, we chose to explore measurement or test apparatus errors, where slew rate, bandwidth, and clipping were the main issues, and not circuit construction errors. So, while previously we focused on faulty apparatus that created a malfunction, this work focuses more on aspects of input signals that push op-amps outside of ideal circuit behavior. In this way, we hope to address different modeling pathways by introducing a distinct context and types of errors.

D. Phase 1 outcomes

Our assessment development process is incremental, with each phase building on the previous. We began by gathering electronics and optics instructor perspectives of modeling in their laboratory courses in phase 1 [9]. The interviews and subsequent analysis yielded how instructors used and valued the modeling process. We also determined that assessments that are aligned with specific contexts (e.g., electronics) need to be developed, since modeling is perceived and used differently in different physics subject areas. For example, electronics instructors often called their courses a "10% science" (in reference to measurement accuracy) or a "yes-no subject" (p. 14), which informed how students evaluate whether a match to prediction is good enough to consider the discrepancy reconciled.

In the phase 1 study, we were also interested in common ways the inverting amplifier configuration is used in circuits, and how students model op-amps and op-amp circuits in electronics labs. This was to ensure that our activity can be used by a diverse set of institutions by addressing contexts and models used widely. During the interviews with the electronics instructors, we focused on lab activities they conducted with inverting amplifier circuits. We learned that passive analog electronic circuit elements, such as resistors and capacitors, were commonly used, as were oscilloscopes, digital multimeters (DMMs), and function generators. Instruction on op-amp models was often based on knowing the "golden rules" of op-amp operation and their implications. That is, the op-amp may be regarded as a black box in which the inputs have infinite input impedance, and will adjust its output so that the voltage difference between the two inputs is zero while under negative feedback. Op-amps were commonly used to create active filters, amplifiers, and rectifiers.

In terms of the physical and measurement apparatus and phenomenon that were commonly used, an important outcome from phase 1 was that "[t]here is no clear evidence for designing an instrument that targets only one particular subtask" of the Modeling Framework [9]—that is, phase 2 should seek details about how students navigate all the subtasks. Coupled to this finding was the instructors' descriptions of the importance of the making comparisons and make measurements subtasks, and their observation that students experience difficulties in the propose causes subtask. Thus, in phase 2 we also seek to enumerate and describe the comparisons, measurements, and proposed causes students may undertake. Further, phase 2 research goals were also informed by the emphasis in phase 1 "that the assessment should elicit information about students' justifications for prioritizing some modeling pathways over others" (p. 19).

IV. METHODS

A. Activity design

Common operational amplifier (op-amp) models treat the integrated chips as "black box" circuit elements [9]. That is, solid-state physical models of the transistors that make up the op-amp are not extensively considered. Practically, it behaves according to gain equations derived from the golden rules, with a frequency dependence that does not follow from the golden rules [31].

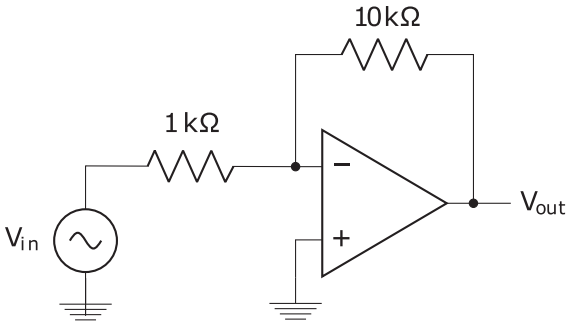

FIG. 2. A circuit diagram of the inverting amplifier circuit used in the TAPS interviews. The students were given a handout with this circuit diagram at the beginning of the activity. The values of the feedback and input resistors were written. Purposefully missing are the values of the power rails (initially set at ±10 V), and the capacitors from each power rail to ground (100 pF)

The central component of our activity was an inverting amplifier circuit (LF356 op-amp) operating outside its recommended voltage and frequency range at a gain of −10 ($A = -10$). We show a circuit diagram is shown in Fig. 2.

The activity was designed such that there were multiple pathways to resolving the issue. Thus, at the onset of the activity, the circuit was powered before the student was instructed to work on it. The frequency (f), input voltage (V_{in}), and power rails ($V_{S\pm}$) were preset to $V_{\text{in}} = 3V_{\text{pp}}$, $f = 250$ kHz, and $V_{S\pm} = \pm 10$ V. We included capacitors ($C = 100$ pF) from the power rails to ground to reduce spontaneous oscillations. Under these conditions, the output signal is clipped (sine wave cutoff at the trough and crest) by the power rails, and slightly distorted due to the slew rate limit. Despite these discrepancies, the signal was still recognizable to a trained eye.

The slew rate limit of the op-amp determines the maximum rate of change of the output voltage, and depends on both the frequency and input voltage amplitude [31]. Operating above the slew rate limit will result in an output voltage waveform that appears closer to a sawtooth than a sine wave, and with a smaller amplitude. A simulation of the output and input signals at the initial conditions and after readjusting the values is shown in Fig. 3.

B. Interview protocol and video data collection

At the start of the interview, we provided the student with a schematic diagram of the circuit, a data sheet for the op-amp, which included a pinout, and a prebuilt functional circuit. We also included the equation for the gain and reminded the students what to expect for the phase based on the gain equation.

The interviewer read a short prompt to the student before they began their work on the activity, including a reminder to think aloud while they worked on the circuit. The exact wording can be seen in the interview protocol reproduced in the Appendix.

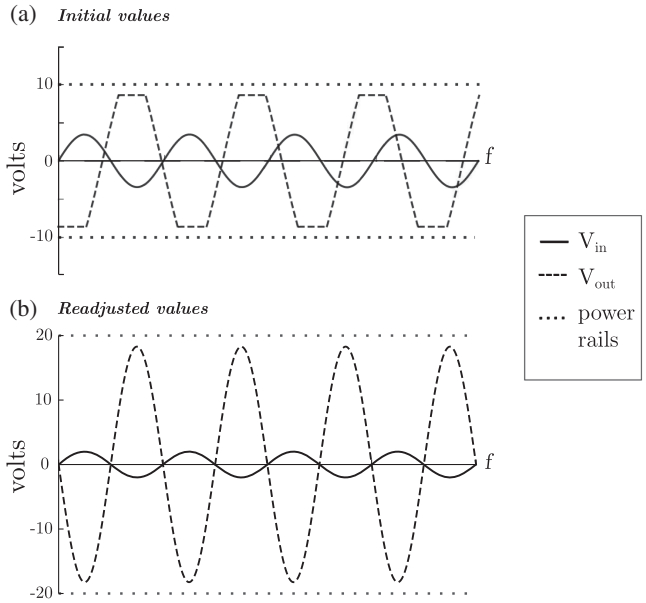

FIG. 3. A simulated plot of the input (solid line) and output (dashed line) voltage signals on an oscilloscope with given power rails (dotted line). (a) With the initial conditions, the output signal will appear clipped due to the power rail limit. The slight slant of the output signal is due to slew-rate limits. (b) After hypothetically increasing the power rails and decreasing V_{in}, the clipping and the slight distortion due to the slew rate limit is fixed.

The interview included up to two problem-solving activities. The first was to make the circuit function as predicted (correct gain, phase, and shape) with a gain of −10, and the second was to do the same with a gain of −100. The second activity was more difficult to complete than the first. We included the second activity to keep the interview going if the student finished quickly. We also wanted to observe if there were differences in modeling between those students for whom the first activity was straightforward and for whom it was not. Of the 10 participants, five declared they had fixed the circuit, and were given the second activity. Of those that reached the second activity, one completed both activities, and four were still engaged in the second activity at the end of the 1 hour interview. Two students did not complete the first activity. Hence, giving the students increasingly difficult tasks after each subsequent hurdle increased the range of the modeling behaviors we observed.

At the end of the TAPS interview, the students were asked a series of questions to gain insight into particular events or modeling tasks that the students might not have articulated aloud, and to understand their academic background in more detail. There were 10 interviews total, all between 55 and 70 min, resulting in about 11 total hours of video data.

C. Participant recruitment and demographics

The participants were junior and senior physics and engineering physics majors who had taken the electronics

TABLE I. Demographic information of participants.

Category[a]	No. of students
Engineering physics major	1
Physics major	9
Third year of study	6
Fourth year of study	2
Fifth year of study	2
Prior research experience	4
Male (Men)	6
Female (Women)	2
White (Caucasian)	8
Did not report race, ethnicity, or gender	2

[a]Since the participants were asked to self-identify their gender and race and/or ethnicity, the terms used in this table for those questions reflect the participants' words.

lab in the physics department (Electronics for Physical Sciences) during Fall 2016, Spring 2017, or Fall 2017. This class is a stand alone, mostly analog electronics laboratory course (see Ref. [25]).

We conducted 10 TAPS interviews during Fall 2017. Students were compensated monetarily for their time. Eight of the ten students answered our final demographic question ("Finally, is it all right if you tell me your gender and your race and/or ethnicity?"). Additional demographic information from the participants was obtained during the postinterview questions, presented in Table I.

D. Coding scheme and data analysis

The *a priori* scheme was developed by using the subtasks (gray boxes in Fig. 1) of the Modeling Framework as the basis for the code categories. L. R. created initial drafts of the coding scheme that were then discussed and revised by the research team.

From the audio-video data, the audio was transcribed. The transcripts were synced with the video data. All of the data were coded together in 30 sec intervals. The interval size was chosen to capture some of the time ordering of subtasks, since revisions and measurements in electronics can be done rapidly in succession.

After the *a priori* coding scheme was created, two authors, L. R. and B. P., completed a training phase before the final coding. The training phase consisted of a collaborative coding segment, followed by an independent coding segment, and reconciliation. After the training phase, L. R. and B. P. separately coded a 10 min segment of the data.

After the first iteration of training and separate coding, we made minor revisions to the wording of the definitions, reconciled differences, and discussed ways in which the robustness of the codebook could be improved. Specifically, we added examples of what to code and what *not* to code. For example, two common revisions were "Changes the scale on the oscilloscope to get a better look at the signal," and "Changes the value of the input voltage, V_{in} on the function generator." Changing the scale on an oscilloscope is a technical skill, whereas changing the value of V_{in} is relevant to the model of circuit behavior. Therefore, we coded changing V_{in} under the enact revisions code, but not changing the oscilloscope scale settings.

After adding these types of examples, we undertook another round of training and separate coding process, achieving a Cohen's kappa of 0.69. In this round, L. R. and B. P. noted the code make measurements had the most disagreements, in particular on how to incorporate students' use of the oscilloscope to make measurements. In electronics, the oscilloscope is constantly measuring circuit performance, so one might interpret glancing at the oscilloscope as a measurement. However, we wanted to code instances where students were articulating measurements rather than just looking at the oscilloscope. With input from all authors, the code definition was amended to capture the use of utterances about observations made on the oscilloscope, reinforced or contextualized by the video data.

The following is an example of the revision to the make measurements code. In this segment, Sexton is articulating an observation about what she sees on the oscilloscope:

> *So now, the gain is only, like 4 again. Yeah, it's not correct.—Sexton*

Before revising the codebook, this instance was not coded under the make measurements code category. After discussion, we reasoned that Sexton was in fact making a measurement and then a statement about the input and output voltages, determined by eye, on the oscilloscope. All other episodes like the example above were coded under the make measurements category.

A subsequent round of training and separate coding yielded an overall Cohen's kappa of 0.88, indicating almost perfect agreement. The improvement also indicated that disagreements in the make measurements code category were affecting the overall score, and that our revision significantly improved the codebook.

L. R. then completed the rest of the coding using the final coding scheme. An entire 30 sec interval would be coded if there was one utterance or sentence that would fit into a code category. Multiple code categories, or no codes, could be assigned to a single 30 sec interval. For example, in the following segment, Auden is measuring the resistor values of the feedback and input resistors with a DMM:

> *This is…1 point 89 (1.89) kilo-ohms. Oops! I mixed up R-f and R-in. This is R-in, that's R-f. Okay, so, that looks pretty okay.—Auden*

The first portion where he measured the resistor values with a DMM fits under the code category make measurements, and his qualitative statement that the measured

TABLE II. *A priori* coding scheme and examples from different student participants.

Code	Definition	Example
Make measurements	Student utters or articulates an observation of the data output from a measurement apparatus or device. In the case of electronics, the measurement device may be the DMM, the oscilloscope, or the student's observations about the circuit construction.	"So the next thing I'm going to do is measure the resistors and make sure they're actually what they say they are."
Construct models	Students use their conceptual models of the measurement or physical apparatus or model to reason through the activity. Model construction can occur as a way to orient about the circuit or activity, or after any other subtask has occurred.	"They're related because of the gain. Because the gain is Vout over Vin, so, I guess if I'm changing Vin, then the output will change."
Make comparisons	Student compares their observed data to a prediction from the model, or expectations from previous work. The comparison may be qualitative or quantitative. The student may also express that the outcome is "good enough."	"They [input and output signals] are definitely between 170 and 180 degrees phase shifted…So I'm going to call that a success."
Propose causes	Student describes or articulates a reason for the discrepancy between measured or observed quantity and prediction. The proposed cause does not have to have a detailed empirical motivation for it to be coded here.	"So the problem is that it's operating outside the region, the operation region."
Enact revisions	Student revises, or changes, a component of the measurement or test (e.g., function generator, power supply) or physical apparatus (e.g., resistors in the circuit). The student may also make changes to the measurement or physical model to resolve discrepancies by adding or changing parameters (e.g., may add an arbitrary offset to the gain equation to account for a dc offset).	"What will happen if I change the gain? […] Right now, I just want to see what'll happen if I switch the resistors."

values fit his expectation ("that looks pretty okay") would be considered a comparison. Thus, this 30-sec interval would be coded in both the make measurements and make comparisons code categories. Conversely, if the student is sitting in silence and not performing actions involving the apparatus for the 30-sec interval, no codes would be assigned. This was not a common occurrence. The final *a priori* coding scheme, with definitions and examples, is shown in Table II.

During the coding process, we cataloged task-specific, emergent subcodes for all code categories. For example, under the make comparisons code category, the act of comparing the expected phase of an inverting amplifying circuit to the observed measurement is a subcode.

At the end of the coding process, some *a priori* subcodes were deemed too sparsely coded, or too specific to be useful. Therefore, some subcodes were collapsed or combined. For example, the two subcodes for measuring V_{in} and V_{out} on the oscilloscope were collapsed into simply, "Voltage measurement on oscilloscope." The finalized emergent subcodes and their relative frequency in the coding are shown in Table III.

V. RESULTS

A. Time spent on individual subtasks

The first research question (RQ1) was meant to capture the most prevalent modeling subtasks in electronics. Within this domain-specific assessment, identifying and then interpreting why certain subtasks are more common will aid in developing relevant test items in phase 3.

To determine the amount of time students spent on a particular subtask, we aggregated the coded data from all 10 TAPS interviews and counted how many 30-sec intervals were coded in the code categories corresponding to the five subtasks in the Modeling Framework: construct models, make measurements, make comparisons, propose causes, and enact revisions.

In Fig. 4, the aggregated data are shown in a schematic highlighting the commonality of codes and their connections. The area of each dark gray circle represents the number of 30-sec intervals that were coded with a particular subtask. The thickness of the gray connectors indicates how many times a subtask was coded directly after another subtask per 30-sec interval, in addition to the number of times each pair was coded in the same 30-sec interval. We use this figure to guide our work in two ways: First, we look at the amount of time spent doing each subtask, which will help in narrowing down the research findings and resulting test items to the most relevant themes within each subtask. Second, the connectors in Fig. 4 indicate which subtasks are clustered together. This will help inform how we examine the processes that students take while modeling, and which processes we expect to observe or evaluate in future phases.

We found that students spent the most amount of time in the make measurements, make comparisons, and enact revisions subtasks. The large area of the make

TABLE III. List of emergent subcodes from the analyzed transcripts. The percentage of coded items refers to how many times a 30-sec interval was coded with each code category (subtask). The last column denotes how many interviews out of 10 had at least one instance of the corresponding subcode.

Code category	Subcode	% coded items	Interviews
Construct models	Relating gain to V_{in} and V_{out}	88	9
	Functional limits of op-amp chip	12	4
Make measurements	Voltage (oscilloscope)	25	10
	Shape (oscilloscope)	17	10
	Visual inspection of circuit	16	10
	Voltage (DMM)	11	6
	Resistor values (DMM)	11	8
	Phase (oscilloscope)	9	9
	"Noise" (oscilloscope)	7	6
	Frequency (oscilloscope)	3	6
	Current (DMM)	1	1
Make comparisons	Compare circuit construction to pin out	22	10
	Quality of waveform	21	10
	Gain is or is not correct	15	8
	Phase is or is not correct	11	10
	Sine shape is or is not acceptable	9	9
	Student articulates that it is "good enough."	8	7
	Predict gain based on values of the resistors	4	7
	Predict phase or shape output	4	6
	Predict overall equipment behavior	3	8
	Predict V_{out} based on gain and V_{in}	2	5
	Predict op-amp circuit behavior	1	2
Propose causes	Values on test or measurement apparatus are too high or low	34	9
	Incorrect circuit construction	30	9
	Capacitance, resistance, or impedance issue	17	8
	Test equipment are faulty	12	5
	Bad op-amp chip	7	4
Enact revisions[a]	Lowers or increases frequency	23	9
	Lowers or increases power rails	22	9
	Lowers or increases V_{in}	20	9
	Switches out resistors	15	7
	Changes function generator settings	7	6
	Re-wires or revises circuitry	6	3
	Changes op-amp chip	5	3
	Switches to scope probes	1	2
	Changes grounds on power supply	1	1

[a]None of the revisions the students enacted were to the model; these revisions correspond only to the measurement or physical apparatus.

measurements subtask may be partly due to the ease with which one may carry out measurements in electronics. Further, we operationalized the make measurements subtask to include utterances about what a student observes on the oscilloscope, which is constantly taking data. However, these potential reasons why make measurements is the most commonly coded item are not artificial; this speaks primarily to how measurements are used and interpreted in electronics. How we will capture the significant role of measurements in the assessment will be discussed in Sec. VI A

Compared to measurements, comparisons, and enacting revisions, students spent much less time proposing causes and constructing models. The lack of items coded in the construct models subtask may be partially explained by how much information we intentionally gave the students about fundamental aspects of the op-amp operation model. The students may not have had much opportunity to construct further models, a tension that will be discussed within the context of subsequent phases of the assessment in Sec. VI A.

B. Temporal connection of subtasks

The second research question, RQ2, seeks to understand which subtasks were clustered together in time to

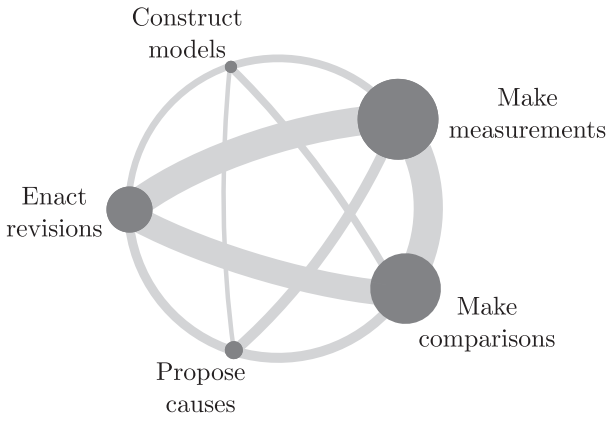

FIG. 4. Aggregated data from participants on duration and sequence of modeling subtasks. This schematic shows the commonality of the five modeling subtasks: construct models, make measurements, make comparisons, propose causes, and enact revisions. The area of the circles represents the number of 30-sec intervals coded with that particular subtask; the thickness of the gray connectors between all five subtasks represents the number of times a subtask was coded directly after the other, or within the same 30-sec interval. This representation is similar to the transition graphs presented in Ref. [32].

identify and describe common, missing, or unexpected pathways. To this end, we counted the number of times a particular subtask in a 30-sec interval (e.g., make measurements) was directly followed by another modeling subtask (e.g., make comparisons). We also counted the number of times the subtasks were coded within the same 30-sec interval (e.g., make measurements and make comparisons). This includes utterances that would be defined as both subtasks simultaneously and subtasks that occur rapidly after one another. The gray connecting lines show the connection between just two subtasks, which we have termed adjacent pairs. The thickness of the connectors denotes how many times an adjacent pair was observed.

From the data, we observe that all adjacent pairs are present. We interpret this as representative of the interconnectivity of the modeling process. While the Modeling Framework describes a recursive, nonlinear process, these data also show that students may undertake the modeling subtasks in any order.

Make measurements ↔ make comparisons, make measurements ↔ enact revisions, and make comparisons ↔ enact revisions are the most prevalent adjacent pairs. This, together with the finding in Sec. V A that these three subtasks are also the three most common subtasks in this dataset, underscores the centrality of iterative measurements and revisions in electronics. When confronted with a challenging issue, students may skip the propose causes subtask in favor of enacting more revisions or taking further measurements (see Ref. [28]).

C. Description of subcodes and themes

1. Make measurements

Students constantly make many measurements of different types. This is noteworthy because it communicates that generally students have facility with many types of measurements related to electronics, and use measurements constantly to inform their reasoning. There were three general types of measurements we cataloged—those accomplished with the oscilloscope, the digital multimeter (DMM), and visual inspection.

Measurements made using the oscilloscope and the DMM constituted the majority of the measurements. Students showed ease with both making the measurements and interpreting the data. For example, the most common measurements where those verifying the voltage, phase, and shape of the input and output signals by visual inspection on the oscilloscope. Sometimes, students made measurements on the quality of the output signal using qualitative descriptors like "noise," "stable," and "shaky." For example, in the following quote, Lorca [33] made successive measurements about the so-called shakiness of the output signal at different input voltages.

> *So now, we got 1 volt input [Vin]. So…[pause] Getting some kind of shakiness at 1 volt from the wave amplifier [high frequency noise pick-up], so I'm going to change it [Vin] to 2 volts. And that seems even more stable.—Lorca*

In the following example, Plath is beginning the activity and has just verified that the circuit construction is correct. She then makes additional measurements on the oscilloscope to measure the values of V_{in} and V_{out}.

> *So, it looks like the…so looks like my peak to peak output is like 5 volts for the input versus…1, 2, 3…I think 3 and a halfish [voltage divisions on the oscilloscope], uh, yeah. So 15, 17.5 volts? So, oh in any case, that's like 3 and a half…a gain of like 3 and a half, which is not 10.—Plath*

Students used the DMM mostly to make measurements on the circuit specifically, if the power supply was functioning correctly and powering portions of the circuit. Most students would also check the resistor values with the DMM.

All 10 participants initiated the activity by visually inspecting their circuit to ensure that it was constructed correctly. Sometimes when they were completely out of ideas, they would revisit the circuit construction as a possible source of error, e.g.,

> *Okay, I'm going to check that everything's plugged in like it's supposed to be, just to double-check. Input is going from the function generator through the resistor*

into…positive input is grounded…and, from the negative one fed back into output…—Frost

Of the subcodes encountered in the make measurements subtask, students struggled with very few of them, indicating an overall adeptness at making measurements. Engaging with *which* measurements are common and conducted with ease, not just looking at the make measurements subtask overall, helps establish guidelines for subsequent phases of the assessment in Sec. VI A.

2. Construct models

Overall, there was not much opportunity for model construction, since the majority of an op-amp model was purposefully given to the students, including the equation $A = -R_f/R_{\text{in}}$. We did not ascribe students calculating gain from the ratio of the resistors to model construction; this would be more accurately described as *using* a model by parametrizing the gain with the given resistor values. Instead, we used two emergent subcodes to describe the models students did construct.

One subcode describes constructing models about the output voltage signal, given the gain and the input voltage, using the equation $A = -R_f/R_{\text{in}}$.

Another subcode encapsulates how students will sometimes construct further models on the relationship between V_{out} and V_{in} to A with the equation $V_{\text{out}} = AV_{\text{in}}$. This model construction process is exemplified by the following quote, where Sexton reasons about how to relate V_{in} and V_{out} directly based on what she measures on the oscilloscope:

Actually, yeah, that makes sense. They're related because of the gain. […] Because the gain is V-out over V-in, so, I guess if I'm changing V-in, then the output will change.—Sexton

The second subcode describes how students construct models of op-amp function when considering the operational limits of op-amp behavior. The model construction around the functional limits of the op-amp included reasoning about frequency, input voltage, and power limitations. Utterances coded here were typically done at the beginning of the activity, likely to assess the initial parameters. In the following example, Lorca constructs a model aloud about how the functional limits of the op-amp will affect the signal he observes:

I guess I understand that whenever something's saturated, it's because you're operating outside the limits of the op-amp.—Lorca

Here, the "saturation" Lorca mentions refers to the clipping of the output signal due to the limited power rails. He is connecting this observation with a practical model of op-amp circuit behavior by qualitatively identifying that op-amps have functional limits.

3. Make comparisons

The make comparisons subtask was about as common as make measurements. The most common type of comparison was checking the circuit construction to the given pinout, data sheet, and circuit diagram. All participants engaged in these comparisons.

We found that the majority of the comparisons were evaluating how well the gain, phase, and shape matched the prediction, aligned with the goals of the activity as provided at the beginning of the interview. In addition, students would compare the general acceptability of the output waveform, including commenting on noise or other unexpected behavior. All of the students articulated that all three output signal characteristics were good enough when they finished the first prompt.

Within the make comparisons subtasks, we noticed that comparisons could either be qualitative or quantitative. Comparisons of shape, phase, and quality would often be qualitative. For example, a student may compare the waveform's qualitative characteristics to some standard of quality and shape:

So it's [the output signal] sort of sinusoidal, but mostly it just appears to be noise.—Frost

In contrast, students often made quantitative comparisons about gain, or constituent voltages, as in the following example from the same student:

Right now the input is at 1 volt peak to peak, and then the output is at 10 volts peak to peak, which is a gain of 10, which is what I want.—Frost

In addition to comparing their data to a model in some way, we also noticed that students would articulate comparisons between successive measurements as getting them closer or further from their goal. They may iteratively compare their measurements as they revise possible parameters. In the following example, Dickinson qualitatively compares the stability of their waveform after enacting a revision to the physical apparatus:

Huh. That actually stabilized the waveform a little bit more. […] I switched the 1 mega-ohm [resistor] with a 1 kilo-ohm resistor, switched the R-f [feedback resistor], made the gain smaller.—Dickinson

Embedded in the operationalization of the make comparison subtask are predictions. The most commonly coded predictions were based on the expected gain from the given values of the resistors, often completed at the beginning of the activity, and at the initiation of the second prompt if the student finished early. Phase and shape predictions were

also common. Many predictions were made about the output voltage given the gain and the input voltage from the function generator. Further, model construction around the V_{out} to V_{in} ratio often led to comparisons using students' predictions about the output voltage amplitude (V_{out} based on gain and V_{in}), and the gain calculated from the resistors.

Interestingly, students would some times also predict that the circuit should be completely functional, but it evidently was not (subcode "Predict overall equipment behavior"). These subcodes are characterized by utterances such as

> *It [the circuit] should be able to produce the gain predicted by the resistances. I shouldn't have to change the…predicted gain by changing the resistors.—Plath*

Here, Plath has just begun the activity and is beginning to reason about the parameters of the gain. She predicts that the absolute resistance values are neither too high or too low, and therefore the gain produced by those resistors is an acceptable value. These predictions played an important role in aiding the student to determine useful measurements or revisions to make if their prediction turned out to be incorrect.

4. Propose causes

The types of proposed causes that the students articulated did not vary significantly between different interviews. Most of the proposed causes were about incorrect values on the measurement or test apparatus, e.g., input voltage that was too high for correct operation of the circuit. An equally common proposed cause was incorrect circuit construction, but that was not a designed flaw in the activity. Other typical proposed causes had to do with faulty equipment, issues with impedance or resistance (e.g., the internal resistance of the test apparatus), or a bad op-amp.

We noticed that many instances of students proposing causes did not exactly address the measurement or comparison the student had just made. That is, a proposed cause may be *connected* or *disconnected*, in reference to whether the uttered proposed cause was connected to the results of the comparison or measurement. We found that often, students would articulate that they understood that *something* was wrong, but did not connect their observations to a specific proposed cause. Indeed, there were several times when students were attempting to understand the nature of a particular discrepancy, and instead of following through with a probable proposed cause, appeared to assign a disconnected cause from a list of typical causes, presumably from experiential knowledge. For example, in the following episode, Frost has maneuvered the circuit into a regime where spontaneous oscillations are obscuring the output signal. When he measures the voltage from the output pin on the op-amp to ground with the DMM, the output signal clears up.

> *So, if I measure from here to here, then it [output waveform] clears up. [measures again, pauses] That is strange. So, I'm trying to figure out what the multimeter does when it's in voltage mode. It is in parallel, and it has a very high resistance. [brief silence] Huh. I'm going to try another op-amp, because I don't…I think it's a different issue.—Frost*

Frost began to reason correctly about how his measurement device, the DMM, would affect his measurement. Instead of proposing a cause having to do with the actual effect of the DMM (resistance of the DMM is in parallel with the circuit board when measuring), he instead opted to propose that there is a "different issue," which would be addressed by changing the op-amp. These episodes exemplify how a proposed cause disconnected from a comparison could be used to inform a revision.

Overall, we found that the propose cause subtask was the area where students appeared to resort to other subtasks when they were unable to initially propose a cause for an observed discrepancy. Most proposed causes were disconnected, highlighting the difficulty around this subtask. We discuss this further in Sec. VI A.

5. Enact revisions

Revisions to any part of the experimental apparatus were the most readily undertaken. Accordingly, the most common revisions were to the test or measurement apparatus: the function generator and the power supply. The power rails were revised frequently because, often, the students did not have a complete model of how these voltage levels would affect the signal, so they would explore the effect of changing these values for a significant amount of time. Changes to the input voltage were similarly exploratory and continuously varying. Instead of trying discrete input voltage values after articulating a proposed effect, they would often search the entire range of voltages available.

Less commonly, the students would revise extraneous settings on the function generator, such as the dc offset and output impedance settings. Only a couple students revised how the measurement or test apparatus was used, such as using oscilloscope probes to change how they took a voltage measurement. Revisions to the physical apparatus were dominated by switching out the resistors. Most of these instances were times when the students reasoned that the initial given values were inappropriate. For example:

> *I should probably use a different ratio of resistors so that um…Maybe smaller? A smaller value, so that both of them are smaller in order to get more current in?*

I don't know. So, before I had 10 kilo-ohms to 1 kilo-ohms, so maybe…1 kilo-ohm to 100 kilo-ohms.—Plath

The students also revised their physical apparatus by rewiring or revising the circuit construction, or by changing the op-amp chip. Several students changed the op-amp chip more than once.

Interestingly, we found no evidence that revisions to either the physical or measurement model were made, even though in principle students could have used a revised equation for the gain which includes a frequency dependence.

Similar to the propose causes subtask, we noticed that the students often enacted revisions that did not stem directly from a proposed cause, and less often, made revisions that directly addressed either a comparison, or a proposed cause. We also called these disconnected or connected, respectively. The differences between the motivations or circumstances behind an enacted revision indicates that the difficulty around enacting revisions is related to general difficulties in proposing causes.

VI. DISCUSSION

A. Implications for design and structure of subsequent assessment phases

In phase 1, we learned that instructors found the make measurements subtask to be one of the most frequently undertaken by students and an important subtask with which to have practical facility [9]. Many spoke about how important it is for students to learn how to use equipment to make measurements (p. 12).

In our data, we find that the make measurements subtask is the modeling subtask that the students undertake most. Additionally, the make measurements subtask forms an adjacent pair with every other subtask. When we analyze the subcodes for the make measurements subtask, we notice that there was a subset of measurements which all students made, mostly using the oscilloscope to evaluate voltage, phase, and waveform measurements. We find that the strong connections between make measurements and other subtasks are partly due to the fact that there are varied reasons why and how a student would undertake a measurement. This observation indicates that measurements dominate the modeling pathway not necessarily just because they are easy to make in electronics, but also because students often need measurements to affirm, discard, or motivate their conclusions.

In short, the make measurements subtask is one that students undertake often, and with ease, and so it is not crucial that we assess student competency with this subtask. However, it is crucial for students to have the opportunity to see the results of measurements to guide their modeling process. Therefore, phases 1 and 2 together demonstrate a need for an assessment that allows students to easily access the results of common measurements.

For phase 3 of the assessment, one way to allow students to make measurements is to include an image of an oscilloscope screen that the students can use to measure the output and input signals. It will also be necessary to include the ability to access the result of a measurement with a DMM. The decision of when and what to measure will be up to the student, but after that decision is made, we will provide the result of the measurement. This is in contrast to allowing students to interact with a virtual measurement instrument like in a PhET simulation [34].

Another outcome from our data is that students often struggle to propose causes; it is the modeling subtask that they spend the least time doing, and, consequently, the least likely to form an adjacent pair with other subtasks. This is not altogether unexpected. In phase 1, both electronics and optics instructors highlighted the fact that their own students seemed to struggle the most with proposing causes for discrepancies. Examining Table III, we see that a couple of the most common types of proposed causes—"Incorrect circuit construction," and "Capacitance, resistance or impedance issue"—are not connected to a discrepancy. Indeed, the circuit was constructed correctly, and many instances of the latter proposed cause were vague or speculative. Thus, it seems that students are selecting from a limited range of concepts for both enacting revisions and proposing causes, and that different students have the same limited set of tools. Both research outcomes—short duration of the propose causes subtask and disconnected revisions and probable causes—highlight the need for assessing the propose causes subtask repeatedly, precisely because this subtask appears to be quite difficult.

We found that the enact revisions subtask was a similarly challenging subtask for students. A majority of instructors from phase 1 identified the enact revisions subtask as one students undertake rarely, citing mostly time constraints and lack of facility with apparatus. Most instructors described instances of students enacting revisions as part of the troubleshooting process [9]. We have previously shown that instructors think troubleshooting is extremely important [35] and that a major goal of electronics is for students to build circuits that work [36]. In addition, instructors believe that since "nothing works the first time" [37] troubleshooting and revisions are a required part of working with electronics.

In the data from phase 2, we find that students also engage frequently with the enact revisions subtask in the context of troubleshooting. The data from phase 2 suggest that we should be deliberate about distinguishing between connected and disconnected revisions, as sometimes students will revise the apparatus as a way to understand the discrepancy before proposing a cause. Therefore, phases 1 and 2 together show that revisions should play a significant role in the assessment and attention should be paid to the subtask's connectedness to a comparison or measurement.

Looking ahead to phase 4, the final assessment should also take into account student reasoning for proposing causes and enacting revisions. We had previously proposed to do this by using a coupled-multiple response assessment for phase 4, where a student is asked to choose a reasoning for their selection in a multiple-choice question (see, e.g., Ref. [38]). The results from this study suggest that the available reasoning elements should include an option along the lines of, "I didn't know what else to do," in order to identify motivations for particular revisions and causes.

Overall, the outcomes of our data analysis from phase 2 complement phase 1 outcomes to create a clearer picture for subsequent phases of the assessment development. As discussed, future phases should probe all five modeling subtasks.

From our analysis of the modeling subtask adjacent pairs here, we affirm the need for a process-oriented assessment that allows us to observe the approaches to modeling students take. To accommodate and observe all approaches, we require a very flexible assessment that would not interrupt the modeling process. However, in Phase 3, we also must probe the reasoning behind the actions and approaches students take in order to create a scoring rubric in Phase 4. One way to achieve both of these goals is to design a two-part assessment with one portion focused on the processes and pathways students take, and the other focused on student competency with each modeling subtask. An assessment with this structure will need to be validated with respect to both individual subtasks and connections between subtasks without interfering with one another. Design and validation of such an assessment is the focus of ongoing work.

B. Limitations of phase 2

Even though we were seeking to show the variety of student behavior around modeling an electronics circuit, we found no evidence that students revised their models of either the measurement or physical systems under investigation. We note that students' ability to construct models is closely dependent on their domain knowledge. This suggests that the final instrument will need to assess domain knowledge to the degree that it allows students to engage in the modeling process. Further, we cannot make any claims about students' revisions of models. This is a significant limitation of this study that we will address by seeking expert physicists' feedback specifically about model revisions in phases 3 and 4 of the assessment development process.

In phase 1 of the assessment, our research team was purposeful about seeking out diverse types of institutions from which to draw instructor interviews. In our vision, the ultimate assessment should be generalizable to various institutional contexts and student populations. Because of time and cost constraints, we were not as broad here in recruitment for participants. Specifically, all the student participants for this study were enrolled at CU Boulder when they were interviewed. CU Boulder is a large, selective R1 (highest research activity), predominantly white institution with a very large physics department [39]. This institutional context introduces limitations to the study, as its results may not translate completely to other student populations and contexts. In contrast, phases 3 and 4 will involve free-response or coupled-multiple response survey data, which are much more conducive to multi-institution studies.

C. Connections to other research directions

While the focus of this research paper was to describe the progress towards an assessment, and pay careful attention to its development and process as suggested by the NRC [1], we certainly foresee that our findings may inform future research on student understanding of electronics.

In our study, we described the centrality of measurements in the modeling process. Russ and Odden have also described the role of collecting evidence to make sense of a system [17], which adds to our description of measurements as a sense making, investigative tool in modeling. We observe similar patterns of student engagement with modeling when we consider measurements as evidence.

We also observed students making continuous measurements and revisions when confronted with a confusing problem for which they could not propose a cause for discrepancy. These episodes varied in time, but often took several minutes during which the student explored the functional limits of the op-amp by making different types of measurements. In this way, this work is distinct from Vonk *et al.* [18] in that the longer timescale of iterative measurements observed in our dataset could illustrate extended instances of model making when students tested the functional limits of the op-amp model. What Vonk calls model making, we may describe using the language of the Modeling Framework as model construction. However, we did not observe model breaking, which we would describe as revisions to the model after critical testing. Instead, revisions were made exclusively to the apparatus. However, Vonk's work suggests model making and model breaking (loosely construed, model construction and model revision, respectively) are necessarily interlocked. In the future, investigating the role of domain knowledge in prompting either model breaking or revisions to the model could shed light on why we did not observe model revisions.

Finally, here we will continue to explore the complexity of students' general difficulty with proposing a probable cause for discrepancy. While left unanalyzed in this work, we note the unexpected existence of propose causes $\leftrightarrow$ construct models adjacent pair. It is interesting to note that some times these two subtasks came together, presumably as a sense making tool. Examining these episodes would be

a valuable way to understand why and how the propose causes subtask is challenging.

VII. CONCLUSION

Our team is currently developing two modeling assessments for upper-division physics electronics and optics laboratory courses in four incremental phases. This work describes the goals and outcomes of phase 2 for the assessment on electronics, and ways our findings productively align with, extend, and complement other work on model-based reasoning. We conducted 10 think-aloud problem-solving interviews with CU Boulder upper-division physics students in order to learn how students engage in model-based reasoning while working on a practicum-style electronics activity consisting of an inverting amplifier circuit. Another goal of this work is to document in detail the assessment development process, per an NRC directive to create evidence-based assessments of experimental physics practices.

Four outcomes of this work are as follows: (i) Students engaged in all modeling subtasks, and they spent the most time measuring, comparing, and revising; (ii) each subtask occurred in close temporal proximity to all other subtasks; (iii) sometimes, students propose causes that do not follow from observed discrepancies; (iv) similarly, students often rely on their experiential knowledge to enact revisions that do not follow from articulated proposed causes.

To accommodate these outcomes, we propose a two-part assessment with a process-oriented section, and a free response survey to assess competency within individual modeling subtasks. We also propose providing the student with the results of measurements and revisions that we observed here to be easily undertaken, and thus, less relevant to assess.

ACKNOWLEDGMENTS

We are grateful to the CU Boulder physics students who volunteered their time to participate in this study. We would like to thank Dr. Kevin L. Van De Bogart for his suggestions on developing the coding scheme, and Dr. Bethany R. Wilcox for her suggestions on improving the manuscript. This material is based on work supported by the National Science Foundation under Grants No. DUE-1611868, No. DUE-1726045, and No. PHY-1734006.

[1] National Research Council, *Adapting to a Changing World: Challenges and Opportunities in Undergraduate Physics Education* (National Academies Press, Washington, DC, 2013).

[2] AAPT Committee on Laboratories, *AAPT Recommendations for the Undergraduate Physics Laboratory Curriculum* (American Association of Physics Teachers, College Park, MD, 2015).

[3] K. N. Quinn, C. Wieman, and N. Holmes, *Proceedings of the 2017 Physics Education Research Conference, Cincinnati, OH* (AIP, New York, 2017), pp. 324–327.

[4] A. Buffler, S. Allie, and F. Lubben, The development of first year physics students' ideas about measurement in terms of point and set paradigms, Int. J. Sci. Educ. **23**, 1137 (2001).

[5] B. R. Wilcox and H. J. Lewandowski, Students' epistemologies about experimental physics: Validating the Colorado Learning Attitudes about Science Survey for experimental physics, Phys. Rev. Phys. Educ. Res. **12**, 010123 (2016).

[6] K. R. Galloway and S. L. Bretz, Development of an assessment tool to measure students' meaningful learning in the undergraduate chemistry laboratory, J. Chem. Educ. **92**, 1149 (2015).

[7] L. A. Corwin, C. Runyon, A. Robinson, and E. L. Dolan, The laboratory course assessment survey: A tool to measure three dimensions of research-course design, CBE Life Sci. Educ. **14**, ar36 (2015).

[8] M. Caballero, D. Dounas-Frazer, H. Lewandowski, and M. Stetzer, Labs are necessary, and we need to invest in them, APS news **7**, 8 (2018).

[9] D. R. Dounas-Frazer, L. Ríos, B. Pollard, J. T. Stanley, and H. J. Lewandowski, preceding Paper, Characterizing lab instructors' self-reported learning goals to inform development of an experimental modeling skills assessment, Phys. Rev. Phys. Educ. Res. **14**, 020118 (2018).

[10] B. R. Wilcox, B. M. Zwickl, R. D. Hobbs, J. M. Aiken, N. M. Welch, and H. J. Lewandowski, Alternative model for administration and analysis of research-based assessments, Phys. Rev. Phys. Educ. Res. **12**, 010139 (2016).

[11] 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).

[12] B. M. Zwickl, N. Finkelstein, and H. J. Lewandowski, Incorporating learning goals about modeling into an upper-division physics laboratory experiment, Am. J. Phys. **82**, 876 (2014).

[13] E. Brewe and V. Sawtelle, Modelling instruction for university physics: Examining the theory in practice, Eur. J. Phys. **39**, 054001 (2018).

[14] I. A. Halloun, *Modeling Theory in Science Education* (Springer Science & Business Media, New York, 2007), Vol. 24.

[15] I. T. Koponen, Models and modelling in physics education: A critical re-analysis of philosophical underpinnings and suggestions for revisions, Sci. Educ. **16**, 751 (2007).

[16] I. T. Koponen and S. Tala, Generative modelling in physics and in physics education: From aspects of research practices to suggestions for education, in *International Handbook of Research in History, Philosophy and Science Teaching*, edited by M. R. Matthews (Springer Netherlands, Dordrecht, 2014), pp. 1143–1169.

[17] R. S. Russ and T. O. B. Odden, Intertwining evidence- and model-based reasoning in physics sensemaking: An example from electrostatics, Phys. Rev. Phys. Educ. Res. **13**, 020105 (2017).

[18] 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).

[19] S. Allie, A. Buffler, B. Campbell, and F. Lubben, First-year physics students' perceptions of the quality of experimental measurements, Int. J. Sci. Educ. **20**, 447 (1998).

[20] F. Lubben, B. Campbell, A. Buffler, and S. Allie, Point and set reasoning in practical science measurement by entering university freshmen, Sci. Educ. **85**, 311 (2001).

[21] A. Buffler, F. Lubben, and B. Ibrahim, The relationship between students' views of the nature of science and their views of the nature of scientific measurement, Int. J. Sci. Educ. **31**, 1137 (2009).

[22] T. S. Volkwyn, S. Allie, A. Buffler, and F. Lubben, Impact of a conventional introductory laboratory course on the understanding of measurement, Phys. Rev. ST Phys. Educ. Res. **4**, 010108 (2008).

[23] B. M. Zwickl, N. Finkelstein, and H. J. Lewandowski, The process of transforming an advanced lab course: Goals, curriculum, and assessments, Am. J. Phys. **81**, 63 (2013).

[24] 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).

[25] 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).

[26] H. J. Lewandowski and N. Finkelstein, *Proceedings of the 2015 Physics Education Research Conference, College Park, MD* (AIP, New York, 2015), pp. 191–194.

[27] D. R. Dounas-Frazer and H. J. Lewandowski, The Modelling Framework for Experimental Physics: description, development, and applications, Eur. J. Phys. **39**, 064005 (2018).

[28] L. Ríos, B. Pollard, D. R. Dounas-Frazer, and H. J. Lewandowski, *Proceedings of the 2018 Physics Education Research Conference,Washington, DC* (AIP, New York, 2018).

[29] K. L. Van De Bogart and M. R. Stetzer, Investigating student understanding of bipolar junction transistor circuits, Phys. Rev. Phys. Educ. Res. **14**, 020121 (2018).

[30] K. L. Van De Bogart, D. R. Dounas-Frazer, H. J. Lewandowski, and M. R. Stetzer, Investigating the role of socially mediated metacognition during collaborative troubleshooting of electric circuits, Phys. Rev. Phys. Educ. Res. **13**, 020116 (2017).

[31] P. Horowitz and W. Hill, *The Art of Electronics* (Cambridge University Press, New York, 1989).

[32] M. P. Čančula, G. Planinšič, and E. Etkina, Analyzing patterns in experts' approaches to solving experimental problems, Am. J. Phys. **83**, 366 (2015).

[33] All names are pseudonyms.

[34] S. McKagan, K. Perkins, M. Dubson, C. Malley, S. Reid, R. LeMaster, and C. Wieman, Developing and researching PhET simulations for teaching quantum mechanics, Am. J. Phys. **76**, 406 (2008).

[35] D. R. Dounas-Frazer and H. J. Lewandowski, Electronics lab instructors' approaches to troubleshooting instruction, Phys. Rev. Phys. Educ. Res. **13**, 010102 (2017).

[36] H. J. Lewandowski, N. Finkelstein, and B. Pollard, in *Proceedings of the 2014 Physics Education Research Conference, Minneapolis, MN* (AIP, New York, 2014), pp. 155–158.

[37] D. Dounas-Frazer and H. J. Lewandowski, *Proceedings of the 2016 Physics Education Research Conference, Sacramento, CA,* (AIP, New York, 2016), pp. 100–103.

[38] B. R. Wilcox and S. J. Pollock, Validation and analysis of the coupled multiple response Colorado upper-division electrostatics diagnostic, Phys. Rev. ST Phys. Educ. Res. **11**, 020130 (2015).

[39] Indiana University Center for Postsecondary Research, *The Carnegie Classification of Institutions of Higher Education*, 2015 ed. (Indiana University Center for Postsecondary Research, Bloomington, IN, 2015).