Evaluating instructional labs' use of deliberate practice to teach critical thinking skills

Emily M. Smith^{1,2} and N. G. Holmes¹¹*Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA*²*Department of Physics, Colorado School of Mines, Golden, Colorado 80401, USA*

(Received 2 July 2019; accepted 21 November 2019; published 4 December 2020)

The goals for lab instruction are receiving critical attention in the physics education community due to multiple reports and research findings. In this paper, we describe a theoretically motivated scheme to evaluate instructional lab curricula and apply that scheme to three implementations of an electricity and magnetism lab curriculum. The scheme has three components: (1) that critical thinking is a context-dependent process for using critical thinking skills to make evidence-based decisions (2) that to make decisions one must have agency and (3) that deliberate practice can be used to effectively teach critical thinking skills. We use this scheme to evaluate the lab instructions for three sets of instructional labs for their use of deliberate practice for teaching critical thinking skills through varied opportunities for students to exercise agency in making decisions about experiments. Our analysis shows that our curricular design did target the experimentation-focused critical thinking skills, but did not strongly align with theoretical recommendations for deliberate practice. The results provide suggestions for improvements to curricular design. For curriculum developers, instructors, and researchers who intend to teach critical thinking in the context of experimental physics, this scheme serves as a tool to create and evaluate lab instructions in terms of the ways in which specific skills are supported through deliberate practice.

DOI: [10.1103/PhysRevPhysEducRes.16.020150](https://doi.org/10.1103/PhysRevPhysEducRes.16.020150)

I. INTRODUCTION

The goals for physics lab instruction at the college level are receiving critical attention in the physics education community due to multiple reports and research findings [1–4]. The American Association of Physics Teachers endorsed a set of learning goals for undergraduate physics labs related to experimental design, modeling, data analysis, communication, technical and practical skills, and knowledge construction [5]. Several research-based lab curricula exist that attend to those goals (see, e.g., Refs. [6–16]). Though some articulate the learning theories driving the curriculum (e.g., the Investigative Science Learning Environments [7]), little research has evaluated whether the instructional decisions align with the learning theories.

In this paper, we describe a scheme to evaluate instructional lab curricula and apply that scheme to two implementations of an electricity and magnetism lab curriculum intended to teach critical thinking and another curriculum intended to reinforce physics content knowledge. The scheme is built from theories about critical thinking,

deliberate practice, and agency and applies them specifically to teaching students to think critically about physics experiments in instructional physics labs.

We have two goals for this paper: (1) describe learning theories to evaluate design decisions for lab curricula that aim to develop students' critical thinking skills in the context of experimental physics and (2) develop and apply a theoretically motivated scheme for analyzing and informing instructional materials provided to students within this context. We apply the scheme to three implementations of electricity and magnetism labs: one aimed to teach relevant physics content knowledge and not to develop students' critical thinking skills (semester 0), one aimed to teach critical thinking skills through physical model testing (semester 1), and one aimed to teach critical thinking skills through physical model building (semester 2). We intend for the scheme to be informative for instructors and researchers who are developing or evaluating lab curricula for teaching critical thinking skills in labs.

II. THEORETICAL BASIS FOR DEVELOPMENT OF THE SCHEME

We first explore three theoretical components for teaching critical thinking skills: (1) critical thinking is a context-dependent process for using critical thinking skills to make evidence-based decisions, (2) to make decisions one must have agency, and (3) deliberate practice can be used to

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

effectively teach critical thinking skills. We use this scheme to conjecture how to use deliberate practice to effectively teach students critical thinking skills in instructional labs. We then characterize the lab instructions according to this scheme and compare to our conjectures.

A. Critical thinking as a context-dependent process that uses decision-making skills

In this paper, we define critical thinking as the evidence-based ways through which we make decisions about what to do and what to trust. We assume that critical thinking (1) involves using and applying critical thinking skills [17] and (2) is context dependent. In this section, we elaborate on and justify this definition and the associated assumptions and describe critical thinking more specifically in the context of experimental physics.

Many definitions for critical thinking have been developed and used in education research (see, e.g., Refs. [19–30]). Though definitions of critical thinking differ, there are some threads that remain consistent, including (1) critical thinking involves drawing conclusions supported by evidence and (2) critical thinking involves making decisions and/or forming beliefs about a situation. Here, we provide several examples to highlight the variability in definitions of critical thinking. Ennis articulated, several times, that “critical thinking is reasonable reflective thinking focused on deciding what to believe or do” (Ref. [27], p. 166). Lipman contrasted critical and ordinary thinking by emphasizing that *critical* thinking requires *criteria*—claims that are backed up by reasons [22]. Kurfiss defined critical thinking as “an investigation whose purpose is to explore a situation, phenomenon, question, or problem to arrive at a hypothesis or conclusion about it that integrates all available information and that can therefore be convincingly open,” and, similar to Lipman, emphasized a two-part aspect: conclusions with justification to support that conclusion (Ref. [24], p. 20). A consensus report for the purposes of educational assessment and instruction summarized “critical thinking to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which judgment is based” (Ref. [23], p. 3).

Prior work has described two components of engaging successfully in critical thinking: skills [19,26,28,30–34] and dispositions [23,25–28,30,35–37]. Critical thinking skills relate to characteristics of one’s cognitive engagement, which is often described by the thinking processes and decision making an individual undergoes, such as formulating questions [20,26,32,34], identifying and examining assumptions [20,26,28,30–32,34], interpreting data [20,23], analyzing data [20,23,30], evaluating data and methods [20,23], drawing inferences from data [20,23,26,31,32,34], developing explanations that support a conclusion [20,23,30,32], and self-regulation [23,26,33]. Critical

thinking dispositions consist of the characteristics that allow or encourage an individual to respond to evidence in productive ways. Such characteristics may include motivation [23,30,36], curiosity [23,36,37], concern for others [27], strategy [36,37], self-correction [26,30,36], and open-mindedness [26,30,37].

In this paper, we assume that written instructions can be designed and structured to help students learn skills but it is more difficult for these instructions to foster dispositions. That is, we assume other forms of instruction (e.g., moves by the instructor, discussions between students, reflection by an individual student) encourage and convey dispositions for critical thinking more efficiently than written lab instructions. Therefore, opportunities for students to learn dispositions may often not be targeted through written instructions, though it may be possible to do so and there may exist curricula where this is the case. Thus, we focus exclusively on critical thinking skills, though future work with other data sources should also evaluate how lab instruction may foster critical thinking dispositions.

The types of critical thinking skills outlined in prior literature apply to many fields (science, technology, engineering, mathematics, and beyond), so critical thinking skills are generally considered to be context independent [38]. However, an individual’s proficiency with applying those skills relies on their content knowledge and the domain or context [22,39]. Discussion about the meaning of “domain” exists in the literature (see, e.g., Ref. [40]); we choose to use “context” because we focus on a specific component of the “physics domain.” People may employ critical thinking skills to ask productive questions when faced with a new problem outside of their expertise, but deciding whether data are valid or the best methodological approach depends on their understanding of the relevant content for a given context. To teach critical thinking in instructional physics labs, therefore, we must further specify critical thinking in the context of experimental physics.

1. Critical thinking skills in experimental physics and lab instruction

Experimental physicists must apply many critical thinking skills and make many decisions throughout an investigation, which have been characterized in different ways by different researchers. We use the Modeling Framework for Experimental Physics to ground the context in which students are expected to critically think [41]. The Modeling Framework for Experimental Physics represents a process through which physicists experimentally develop, use, and refine models of physical and measurement systems. The Modeling Framework assumes that physicists engage in modeling by evaluating and making decisions about models and evidence, and may choose to refine measurement and physical models through acting on comparisons between measurements and models. Building and refining models is an iterative and reflective process that requires

input from and evaluation of measurement and physical models. In this paper, we refer to the processes involved with modeling the physical system as related to the *physical* model and the processes of designing, refining, and evaluating measurements as related to the *measurement* model.

We map the Modeling Framework onto our definition of critical thinking as follows: (1) decisions related to developing, comparing, and improving physical and measurement models tie to *what to do*, and (2) justifications for those decisions tie to *what to trust*. We also synthesize components of the Modeling Framework to identify overarching categories of critical thinking skills in experimental physics: decisions about models of physical systems, decisions about models of measurement systems, decisions about comparing data and models, and decisions about refining models of physical and measurement systems [41]. Within each overarching category, we identify a series of critical thinking skills that correspond to an experimenter's related decisions and actions (see Table I for the finalized list and Sec. III A for the process of developing the list).

2. Critical thinking skills versus other scientific thinking skills

In this paper, we focus on critical thinking skills as a subset of the possible goals for lab instruction, which may be developing content knowledge, scientific skills, or understandings of the nature of science [42–44]. More recently, science education has made a shift toward developing scientific practice [45], which is made up of both successfully executing scientific behaviors (i.e., skills) and evaluating a situation to decide whether to execute particular behaviors (i.e., dispositions) [46,47]. The evaluative elements of engaging in scientific practice are most similar to critical thinking: “Participation in scientific practice by enacting a performance appropriately is based on an ability to evaluate and critique that performance according to how well it meshes with those it supports and according to how well it serves the aim of explaining nature” (Ref. [47], p. 1045). When we refer to critical thinking skills in this paper, we refer specifically to those that facilitate being critical of an experiment or a model, as a subset

TABLE I. Descriptions of critical thinking skills codes with available agency. Only one level of available agency is listed for simplicity. “Cued with examples” is similar to “cued without examples” except that examples are provided in the instructions. “Unprompted” is a null code that refers to a complete lack of the code in the instructions.

Critical thinking skill		Available agency (one of four levels)
Category	Code	Cued without examples
Physical	Design or build	Instructions prompt the development of a physical model(s).
	Justify design or build	Instructions cue students to provide justification for why a physical model(s) is (or is not) appropriate for the situation.
	Predictions	Instructions cue students to use a physical model(s) to make a prediction about expected results.
Measurement	Design or build	Instructions cue students to develop procedures to obtain measurement(s).
	Justify design or build	Instructions cue students to provide justification for their procedures to obtain measurement(s).
	Evaluate data	Instructions cue students to evaluate the reasonableness and/or usefulness of their measurement(s).
Compare	Make comparison	Instructions cue students to make a comparison between physical model(s) and measurement(s) but do not provide a specific method or an example of how to approach the comparison.
	Justify comparison method	Instructions cue students to justify use of a method to compare physical model(s) and measurement(s).
	Evaluate comparison	Instructions cue students to evaluate the comparison of physical model(s) and measurement(s) without providing specific descriptions of how to do so.
Refine	Design refined physical model	Instructions cue students to design or propose a refinement to a physical model(s) based on comparison of physical model(s) and measurement(s).
	Justify refinements to physical model	Instructions cue students to explain why a particular refinement to a physical model(s) is (or is not) appropriate.
	Design refined measurements	Instructions cue students to design or propose refinements to method and/or ways to improve measurement(s).
	Justify refinements to measurements	Instructions cue students to explain why they chose particular steps to refine and/or improve measurement(s).
Cycle	Cycle	Instructions cue students to cycle through experimentation processes without mentioning specific critical thinking skills.

of higher-order cognitive skills used in science [48]. Of the decisions and actions listed in Table I, those related to evaluations, refinements, and justifications are particularly necessary for critical thinking. Building models, designing procedures, and making comparisons may also be classified as general scientific abilities [49], procedural goals [44], or scientific thinking skills [50]. Critical thinking is such that students “are encouraged to: a) ask the question how the model could be improved; b) develop an evidence-based argument to justify their claims” (Ref. [51], p. 54), which particularly draws on evaluating, refining, and justifying.

B. Agency to practice critical thinking skills

With critical thinking skills defined to be related to making and justifying decisions, there is a clear link to a critical thinker having agency. We work from a definition that an *agent* is someone who makes decisions to achieve a goal [52]. We distinguish an agent from an actor, who is carrying out procedures without necessarily making the decisions themselves [53]. We also recognize that an agent may or may not be autonomous, as an *autonomous* agent has free will to make their own choices about what to do and think [54]. In an instructional setting designed to teach critical thinking skills, students are unlikely to be autonomous agents because the instructor has imposed value and constraints on what they expect learners to do and think.

An agent must make decisions to achieve a goal and have motivation to achieve that goal. These two characteristics of an agent may be mapped onto the skills and dispositions of critical thinking. The decisions an agent makes to achieve a goal include setting goals, planning how to achieve them, carrying out the plan with self-regulation, and reflecting on whether the goals were sufficiently met [52,55,56]; these decisions require many critical thinking skills. However, the literature also describes that the agent must have the motivation to move through this series of decisions [52,57] and a sense of self-efficacy that they can achieve their goal [58,59], which require critical thinking dispositions.

C. Teaching critical thinking skills through deliberate practice

The best practices for teaching critical thinking skills are unclear and researchers are divided along two perspectives: general or specific approaches [40]. Generalists advocate for teaching critical thinking as a skill across multiple domains, often within dedicated activities separated from a specific domain. Variations on the separation include a dedicated critical thinking skills course separated entirely from a specific domain [40] to dedicated critical thinking units or lessons within a course on the domain [40]. Alternatively, specificists support critical thinking instruction embedded within the domain [18,40]. Within specific approaches, there are several variations including *infusion* and *immersion* of critical thinking skills [40]. Infusion approaches make these skills explicit to the learner within

the domain. Immersion approaches require that learners engage deeply with the subject matter to use and apply the skills; however, the skills are not explicitly introduced or discussed.

Meta-analyses of literature on these methods for teaching critical thinking have presented conflicting results [38,60,61]. Findings of Abrami *et al.* suggest that students develop critical thinking skills most effectively through the specific approach of infusion or a mixed approach (where critical thinking is taught using a general approach but within a specific context) [60,61]. However, the findings from Tiruneh *et al.* suggest that general or mixed approaches in higher education are more effective [38] though they discuss that low numbers of studies, characteristics of instrument items, and differing instructional strategies may have biased their results.

With our focus on teaching critical thinking *skills* specifically, we also use literature for how learners acquire skills generally. We expand from the claim that learning a skill requires deliberate practice: repeated and targeted practice with that skill in response to feedback from an expert [62,63]. Deliberate practice requires that tasks are developed to attend to and improve the learner’s overall performance at that skill. Applying deliberate practice to learning critical thinking skills “is not just thinking critically about some topic ... It also involves doing special exercises whose main point is to improve critical thinking skills themselves” (Ref. [64], p. 43). Ericsson *et al.* contrasts deliberate practice with “play” to emphasize that deliberate practice is highly structured, requires high effort, and may not be enjoyable [62].

Schwartz *et al.* claim instructional materials that support deliberate practice include the following: (1) a selection of tasks that appropriately challenge the learner, (2) feedback to help the learner refine and improve their performance, (3) tasks that allow for focused concentration during and rest between practice sessions, and (4) motivation for improvement, e.g., incentivized deliberate practice [65]. Tasks that appropriately challenge the learner must be challenging enough that the learner does not immediately know how to complete the task but not so challenging that the task is not yet possible for the learner [66], similar to Vygotsky’s zone of proximal development [67]. For a learner to improve, targeted feedback is required to refine the skill and the learner must engage in effortful and focused practice. The teacher is responsible for identifying what the learner needs to improve and designs a task to specifically address that goal, providing increasingly more challenging tasks as each task is mastered [66,68]. The type and level of the feedback and subsequent practice, therefore, may not be determined *a priori*. As such, there are no predefined rules for the amount of time spent on any single task or the level of difficulty over time [66]. Finally, there must be motivation for the learner to improve, which may be intrinsic or extrinsic.

Deliberate practice, however, is a relatively structured activity, which may seem at odds with providing the agency necessary for engaging in critical thinking. Providing agency does not necessarily mean eliminating structure [69,70], particularly when the goal is to teach students how to make good decisions. That is, for a student to *use* critical thinking skills, they must have agency such that they can make their own decisions. For a student to *learn* critical thinking skills, however, they cannot be entirely autonomous and instructions must support them as they make decisions. There is a spectrum for the possible agency made available to students through the number of and types of decisions they can make [71,72]. Previous work found that students take on more decision making in labs with more structure, as long as the structure supports agency by cuing students for decisions that need to be made [70].

Consider the two sets of lab instructions in Fig. 1 associated with measuring the voltage difference across and current through a resistor in series with a power supply. In both cases, the students are expected to make the same measurements with the same equipment, but the students in the second case have to decide how they will connect the circuit, how many data points they will take, and through what ranges of voltage and current they will span. Some decisions are not available to the students in either case, such as that they should use digital multimeters to make their measurements. The set of instructions in the second case has less structure (providing students with more agency), but the structure has not been entirely eliminated.

Some research-based lab activities and curricula are designed to provide students with agency to make decisions

about different aspects of the experimentation process such as drawing conclusions from data, experimental design, and appropriate experimental analyses [7,11,73–78]. Research on students' learning in these labs show improvements in data analysis skills, understanding of measurement, and quality of experimentation decisions [6,49,77–79]. Previous work, however, indicates that several weeks are required for students to develop specific scientific abilities [80], which suggests that deliberate practice is needed to gain expertise in experimentation. However, most research that has focused on students' decision-making or critical thinking skills has not specifically analyzed the structure of lab instructions provided to students through a deliberate practice lens.

D. Examples of how the scheme could distinguish instructional choices in written lab instructions

From the literature, we formed possible cases for how written lab instructions support (or do not support) development of critical thinking skills. The cases are formed based on two characteristics of the lab instructions: how often a critical thinking skill is explicitly mentioned and the extent to which agency is available for that skill.

To integrate deliberate practice in an instructional lab, we assume the critical thinking skills must be practiced across multiple lab activities and in different contexts. A lack of repetition across different lab activities demonstrates that deliberate practice for a particular critical thinking skill is not supported. Repeated presence of statements associated with a particular critical thinking skill on sequential lab instructions is indicative of targeted and regular practice of that critical thinking skill. Repetition and eventual fading of statements for the critical thinking skill in the instructions implies that, once faded, the students are expected to use that skill on their own. In a one-on-one tutoring context, deliberate practice would be incorporated such that feedback and practice tasks are tailored to each individual learner. However, the written instructions for instructional labs are developed with assumptions about what may be challenging to the learner but are not directly responding to the learner's needs.

The amount of available agency while engaging in a critical thinking skill corresponds to the amount of structure provided for students to engage in that skill. Students would have no agency for making a particular decision if the instructions tell the students what to do (e.g., compare data to a model using weighted least-squares fitting). Students would have ample available agency for making a particular decision if the instructions do not mention the activity at all. In this work, we consider group-level agency because lab instructions are intended to be referenced by a group as they make decisions in the lab.

Four different cases, outlined below, were developed prior to conducting the analyses and are not exhaustive. We discuss *post hoc* interpretations in Sec. IV.

- (1) "For the ammeter, use your tan digital multimeter (DMM) set to 20 mA on the DCA scale. Plug a red wire into its positive 'mA' terminal and a black wire into its negative COM terminal. Then connect the red wire to the positive terminal of the power supply and the negative terminal such that the Ammeter is in series with the resistor. For the voltmeter, use your orange DMM set to 20 V on its DCV scale. Plug a red wire into its positive 'V' terminal and a black wire into its negative COM terminal. Then connect these wires across the resistor while observing the proper polarity. Raise the voltage slowly from zero, taking a dozen readings of the voltage difference across the resistor and the current through the resistor in the range 0 to 15 V. Record your readings in the table below.";

(2) "Work with your group to use the digital multimeter (DMM) to measure the voltage difference across the resistor and the current through the resistor. Describe how you made the measurements in your lab notes."

FIG. 1. Examples of lab instructions targeting the same decisions but with differing agency available to students. Example (1) provides students with precise steps, leaving no decisions to the students. Example (2) requires that students develop a process to make the measurements.

1. Case I: Lab instructions are not structured to involve students in critical thinking skills associated with experimentation

If lab instructions are not structured to involve students in any skills associated with critical thinking in the context of experimentation, then we expect that the coding scheme would rarely be appropriate to apply to the instructions. That is, no codes would apply to the instructions.

2. Case II: Lab instructions provide students with no or little available agency to practice critical thinking skills

Lacking available agency could take two forms in lab instructions. If there was no agency available to students for a specific critical thinking skill, then we expect most decisions surrounding that critical thinking skill to be provided to students. A more severe limitation on available agency would be that all decisions are provided to students, leaving no space for students to practice critical thinking skills.

Alternatively, the lab instructions may never explicitly provide or cue for a skill, which may not provide students with enough structure to be aware of the agency available to them. Effectively, this may be perceived as no available agency for that critical thinking skill, even though, technically, all decisions are available to them.

3. Case III: Lab instructions do not engage students in deliberate practice with critical thinking skills

A lack of deliberate practice could manifest in several ways. First are instances where students are expected to perform a skill once and then spontaneously apply or use the skill in appropriate contexts. Instructions that include these instances are identified by cuing students to use a critical thinking skill once throughout a semester and never cuing that skill again. *Deliberate* practice requires repeated, targeted practice with feedback, which suggests that a critical thinking skill needs to be cued multiple times throughout a semester.

However, merely repeating cuing for a skill is also not sufficient for deliberate practice. In “exercises” students repeatedly practice a skill [81]; however, the structure surrounding that skill remains consistent. Without the opportunity to spontaneously apply or use the skill in new contexts, the skill may become an algorithmic response to a cue rather than a critical thinking skill. Instructions that persistently cue students to use a critical thinking skill without fading the structure or frequency of the cue over a semester are not expected to engage students in deliberate practice.

4. Case IV: Lab instructions use deliberate practice to teach critical thinking skills

Figure 2 provides a possible visual representation of the use of deliberate practice to develop a critical thinking skill.

FIG. 2. Example of deliberate practice to develop a critical thinking skill (case IV). Shading represents the amount of available agency for practicing the critical thinking skill; lighter color means more agency available to groups. Each box represents lab instructions for a single lab activity and time goes from left to right.

In the figure, each box represents a single lab activity, time runs left to right, and the shading of the box represents the amount of available agency. When the skill is introduced, there is less agency (more structure, shown by darker shading) for that skill such as providing the decision for the students. Throughout the semester, structure for the skill is faded (increased agency, shown by lighter shading) so that students have more agency and opportunities to apply the skill in new contexts. However, there is repeated cuing so that students engage in deliberate practice, possibly with feedback from their experimentation processes or from the instructor. Eventually, all cuing for the skill is faded from the lab instructions (no shading in the box representing the instructions for the final lab activity) so that students have full agency, are aware of the skill, and have engaged in deliberate practice with that skill.

III. METHODS

We evaluated the lab instructions for three implementations of the same introductory electricity and magnetism lab course. These courses typically enroll students who are prospective physics majors and minors. To specify our language, “lab instructions” refer to the handouts that were prepared by the curriculum designers and provided to students. We use “lab activity” to mean a single unit of investigation that spans either one or two lab sessions. Each set of lab instructions correspond to each lab activity; Fig. 3 provides the names of the lab activities that we analyzed.

The lab instructions for the first implementation (semester 0) were designed to reinforce students’ understanding of concepts from lecture. These materials corresponded to the traditional labs used in the department. The materials from the other two implementations (semesters 1 and 2) were designed by the authors deliberately to teach a narrow set of critical thinking skills. The lab activities in semesters 1 and 2 were not designed to align with the course content but were embedded within the disciplinary content (electricity and magnetism). The overarching learning goals for the lab sequence were as follows (see Ref. [82] for an overview of how the learning goals were designed and activities were aligned to those goals).

1. Collect data and revise the experimental procedure iteratively, reflectively, and responsively,
2. Evaluate the process and outcomes of an experiment quantitatively and qualitatively,

Lab Week	Semester 0	Semester 1	Semester 2
1	1. Electrostatics*	1. Electrostatics*	1. Modeling Electrostatics Phenomena*
2	2. The Oscilloscope	2. Ohm's Law**	2. Circuits**
3	3. DC Electric Circuits**		
4	4. Kirchhoff's Electric Circuit Rules	3. RC Circuits	3. Electricity! Magnetism! Fun!†
5	5. Capacitors	4. Earth's Magnetic Field ⁺	
6	6. Magnetic Fields and Forces ⁺	5. RC & LR Circuits	4. Magnetic Fields and Coils ⁺
7	7. Inductance & Electrical Oscillations		
8	8. Magnetic Induction	6. Faraday's Law [‡]	5. What does this thing do?
9	—		

FIG. 3. Order of lab activities for semesters 0, 1, and 2. Titles reflect what was on the instructions provided to students. Symbols and colors indicate “matched” lab activities that use similar equipment and investigate similar physics phenomena.

3. Extend the scope of an investigation whether or not results come out as expected,
4. Communicate the process and outcomes of an experiment, and
5. Conduct an experiment collaboratively and ethically.

In response to the semester 1 lab implementation, we developed semester 2 instructions to attend to perceived limitations. We made informal observations of the success of the implementation through in-lab observations, student feedback, and looking through students' group lab notes. Students often seemed to assume there was a “correct” model (either one they had in mind or one provided by the lab instructions) and tended to end their investigation after confirming the model rather than critically evaluating model limitations or following up on unexpected results. Therefore, the lab instructions for semester 2 were redeveloped to more prominently feature phenomena that were not introduced in other parts of the course (e.g., light emitting diodes) so that students could authentically generate explanations and physical models of the phenomena through experimentation. We intentionally shifted the focus away from physical model testing as in semester 1 to a focus on model building in semester 2.

Between these three semesters, we expected semester 0 to provide a baseline comparison and serve as a validity check on the applicability of the scheme to written lab instructions. We expected semester 0 to align more strongly with cases I or II in the previous section and semesters 1 and 2 to align more strongly to cases III or IV.

A. Development of coding scheme

We followed Chi's steps for quantifying qualitative data: selecting and limiting the content to be analyzed, developing the coding scheme, clarifying evidence from the content to refine the coding scheme, mapping the scheme to the data, developing visualizations of the mapping, and seeking and interpreting patterns in the mappings [83]. We focused on analyzing the lab instructions (as opposed to video of in-lab observations or student products) as a first step in developing the methods to evaluate instruction for teaching critical thinking skills. The written lab instructions represent the instructional materials that were most

deliberately constructed by the curriculum developers, and thus best supports the evaluation of the instructional intent, separate from the success of the implementation [84,85]. The instructor's moves in the classroom and the students' behaviors or written products are applications or derivatives of the curriculum design. The goal in this work is to evaluate the theoretical motivations of the curriculum design, and so the written text in the lab instructions is the best at doing so. This narrow focus also allows us to distinguish instruction on critical thinking skills separate from supporting student disposition, though there are likely aspects of the instructions that support dispositions.

We then conducted content analysis on the lab instructions in all three semesters. We developed a coding scheme (Table I) after all three semesters were fully implemented. We (as curriculum developers) designed the lab activities for semester 1 and semester 2 based on research about how students learn, including implicit—but not explicit—forms of the learning theories that we describe in this paper. We (as researchers) began to explicitly view the design of the lab curriculum in the language of these learning theories only after the lab activities were implemented. Through content analysis methods, therefore, we identified the theoretical constructs before analyzing the text [86].

To develop the list of critical thinking skills (Table I), we generated a list of decisions and actions that students make while carrying out experiments aligned with the Modeling Framework for Experimental Physics [41] and our definition of critical thinking. Some decisions in the Modeling Framework were not included in our scheme because they did not directly relate to critical thinking skills. For example, we excluded the action of making raw measurements because critical thinking is mostly involved in the design of those measurements or interpretation of the data after. Separate from design and interpretation, the collection of data is procedural.

We also characterized the level of agency available to student groups for each critical thinking skill:

1. No agency: Specific decisions or steps for the critical thinking skill are provided to student groups so that there are no decisions available to them.

2. Cued with examples: There is a cue for the critical thinking skill with an example or examples of decisions to be made. The group may use the example or develop a different response to the cue.
3. Cued without examples: There is a cue for the critical thinking skill; however, there are no examples provided.
4. Full agency: There are no cues for the critical thinking skill, which leaves student groups to decide whether and how to use the skill.

Individual statements targeting the critical thinking skills were coded for levels of agency as none, cued with examples, or cued without examples. No statements were coded for full agency, as the absence of coded statements represents full agency. A few whole-class discussions were referred to in the lab instructions. We chose to characterize these as “cued with examples” because, at the group level, an individual group is likely to receive examples from another group. Whereas, at the class level, the whole class would instead have agency that is “cued without examples.”

As a first pass, we applied the codes to the lab instructions for a mechanics lab activity with the same overarching learning goals as the labs developed for semesters 1 and 2. We then refined and adapted the codes for clarification and to ease application of the codes to lab instructions.

We then applied the codes to the lab instructions for one activity analyzed for this paper to check applicability. We

then made minor refinements to further clarify the codes. After the codes were refined, we each coded all lab instructions included in this analysis. We compared our independent coding of the instructions and came to full agreement through discussion.

Descriptions of the codes are provided in Table I for the available agency level of cued without examples. We provide excerpts from the lab instructions for the electrostatics activity and for the Electricity! Magnetism! Fun! activity and the corresponding codes in Tables II and III. We chose these activities because they both underwent few changes between implementations, though the timing and names changed.

B. Representation of codes

We used the representation developed for an individual skill in Fig. 2 to represent each coded statement across all activities each semester. We used colors to distinguish categories of critical thinking skills (physical, measurement, compare, refine, cycle) and shades to distinguish the level of available agency (none, cued with examples, cued without examples, full). We then collapsed codes for each critical thinking skill by lab activity. We selected the amount of available agency for each skill by the “minimum” available agency, i.e., by identifying the statement for each critical thinking skill that was coded with the least available agency. We selected to use the minimum because

TABLE II. Application of coding scheme for an excerpt from the Electrostatics activity in semester 1.

Lab instructions	Critical thinking		Available agency
	Category	Skill	
Your instructor will show your group a demonstration. After the demonstration, brainstorm with your group to develop several possible, testable models that could be sufficient or insufficient in explaining the demonstration.	Physical	Design or build	Cued without example
Document as many ideas as you can come up with in your lab notebook. Include descriptions of how you would test the explanation and what you should observe if the explanation holds.	Measurement	Design or build	Cued without example
	Physical	Predictions	Cued without example
For each model, carry out your experiment and carefully document the outcomes.			
Does the evidence support or refute the model?	Compare	Evaluate comparison	Cued without example
If all your models are refuted, develop a new model based on the evidence that you have gathered.	Refine	Design physical model	Cued without example
If one (or more!) of your models are supported by evidence, design and conduct additional tests that will probe the limits of the model. For example, can you extend the model by testing new variables (e.g., different clothes or rods)?	Refine	Design measurement	Cued with example
Continue this process until you have ample evidence to support one (or more!) model using your demonstration equipment.	Cycle		

TABLE III. Application of coding scheme for an excerpt from the Electricity! Magnetism! Fun! activity in semester 2.

Lab instructions	Critical thinking		Available agency
	Category	Skill	
After your instructor assigns your group a variable to test, devise a plan to qualitatively and quantitatively test the effects of that variable.	Measurement	Design or build	Cued without example
◦ You may use any equipment or measuring devices that are available in the lab (that you know how to use) and any additional equipment that has been brought in for this lab.	Measurement	Design or build	Cued with example
◦ Build a body of evidence that rules out possibilities and provides support for other possibilities.	Physical	Design or build	Cued without example
◦ Propose and evaluate several models that may mathematically explain how the variable relates to measurements your group is making.	Compare	Evaluate comparison	Cued without example
◦ Use qualitative evidence to propose several possible models and quantitatively compare your data to each model.	Physical	Design or build	Cued without example
	Compare	Make comparison	Cued without example

At the end of the first week, your group will quickly report out to the class what you have found about the effects of the variable you're testing.

the statements that provided no agency were central to using the particular critical thinking skill in that activity. Using the median available agency tended to overrepresent the amount of agency groups had available to them throughout the activity. Because full agency is only inferred through absence of statements, the maximum available agency is always full and, therefore, not useful. The collapsed representation deliberately does not capture the variations in available agency within an activity. It is also not normalized by the number of statements that were

coded, so we provide a count for the number of statements coded for each critical thinking skill.

The full set of collapsed codes for all three semesters are shown in Fig. 4. Figures 5 and 6 show the codes corresponding to each individual statement in the order they appeared in the lab instructions for each activity in semesters 1 and 2. Many statements were not coded and, therefore, are not represented in the figures. The coding for semester 0 is not represented this way because there are many more coded statements

FIG. 4. Collapsed representation of coding for the lab instructions for each activity; one activity is a single column. Dark shades represent less available agency and lighter shades represent more available agency; the statement coded with the least available agency was used to determine the shade. The numbers represent the number of times statements were coded within the lab instructions for that activity. Activities are ordered by timing from first of the semester (left) to the final activity of the semester (right).

FIG. 5. Codes applied to semester 1 lab instructions ordered by the appearance of the code in the lab instructions, and lab activities are sequentially ordered. Colors correspond to the category of the code and shade provides the level of agency (less available agency identified by a darker shade).

FIG. 6. Codes applied to semester 2 lab instructions ordered by the appearance of the code in the lab instructions, and lab activities are sequentially ordered. Colors correspond to the category of the code and shade provides the level of agency (less available agency identified by a darker shade).

to the extent that this representation is uninformative. In Sec. IV, we draw from all three figures to interpret the degree to which deliberate practice was implemented through the number of statements coded for the critical thinking skills and the corresponding patterns of agency.

IV. RESULTS

Figures 4–6 show the individual codes and the collapsed codes, respectively, applied to the lab instructions. In the following sections, we interpret the figures according to the components of the scheme.

A. Do the lab instructions involve critical thinking skills in the context of experimentation?

Many critical thinking skills were absent in the instructions for semester 0, particularly in the refine and cycle categories. We, therefore, claim that semester 0 instructions align closely with case I for the refine and cycle categories. This result aligns with the aim of semester 0, which was not designed to teach critical thinking skills. Many statements in the physical and measurement categories were coded, but with little or no available agency, which we will discuss further in Sec. IV B.

In contrast to semester 0, the lab instructions for semesters 1 and 2 include statements that target most of the critical thinking skills, suggesting the lab activities generally focused on teaching critical thinking skills. As shown in Fig. 4, each lab activity (represented by a single column) each semester had statements in the instructions coded for several critical thinking skills. Five activities—four in semester 1 and one in semester 2—have statements coded across all five categories of critical thinking skills; four additional activities—one in semester 1 and three in semester 2—include statements coded in all categories except cycle.

Within the physical category, the design or build skill appears in instructions for all lab activities for both semesters. However, statements coded as justify design or build and predictions are less frequent. From semester 1 to semester 2, a justification statement was removed from an activity early in the semester. Predictions statements appeared in one additional activity in semester 2.

For codes categorized as measurements, design or build statements were in all lab instructions for both semesters. Justification statements were in nearly every lab activity in semester 2, but only in the two earliest lab activities in semester 1. An evaluate data statement appeared in only one activity in semester 1 and never in semester 2, which suggests this critical thinking skill was relatively neglected from our design of the lab instructions.

For the compare category, statements coded as make comparison and evaluate comparison each appeared in five out of six lab activities in semester 1, but only three out of five activities in semester 2. The frequency that statements

were coded as justify comparison was also reduced between semesters, with only one such coded statement in semester 2.

In both semesters 1 and 2, most of the statements coded in the refine category were coded as design refined physical model or design refined measurement, rather than justifying the refinements. This aligns with the more frequent inclusion of statements coded as design or build compared to those coded as justify design or build in the physical and measurement categories. In fact, there were no statements coded as justify refined physical model in either semester, suggesting that this skill was not integrated into the lab instructions. There were also some differences between semesters. Semester 2 had more activities that included statements coded as design refined physical model and semester 1 had more activities that included statements coded as design refined measurements.

Cycle was coded frequently in lab instructions during semester 1; however, it was only coded in the first activity of semester 2. With a reduction in the number of statements categorized as refine in semester 2, it is surprising that statements falling into the more ambiguous category of cycle were also removed.

Overall, semester 2 lab instructions tended to more frequently include statements that targeted justify design or build for measurements but reduced the number of statements that targeted most other skills. Statements coded as design or build in the physical and measurement categories appeared in all lab instructions in all three semesters.

B. How do the lab instructions provide available agency to student groups?

Lab instructions for semester 0 provided very little available agency to students, as indicated by the darker shading throughout Fig. 4. Most notably, statements coded as design or build in the physical and measurement categories overwhelmingly provided no agency, meaning many (and usually most) related decisions were provided to the student groups. The absence of statements coded in the refine category could indicate full agency; however, as described in case II, students may not be aware that decisions are available to them in refining the measurement or physical models. This analysis supports the interpretation that, overall, semester 0 strongly aligns with case II: The lab instructions provided student groups with little to no available agency to deliberately practice critical thinking skills.

One interesting outlier is in the comparison category for semester 0. There was agency available to student groups for making and evaluating comparisons. While semesters 1 and 2 included embedded activities to develop several statistical methods for making comparisons, no such statistical development occurred during semester 0. Nevertheless, the instructions occasionally prompted groups to make and evaluate comparisons without indicating how to do so. With no instructional support, this does not align with deliberate practice.

Comparing semesters 1 and 2 overall, the instructions in semester 1 provided less agency. Figure 6 shows the increased available agency in semester 2 by the removal of a substantial number of statements (i.e., fewer statements were coded), and Fig. 4 demonstrates the increased available agency on “minimum” for the activities (i.e., the darkest shades, and thus statements with the least agency, in semester 1 become lighter shades and thus increased agency in semester 2).

In semester 1, there was particularly little available agency in the physical category; that is, groups were often given a physical model (shown by dark red in Figs. 5 and 4) so there was no available agency for them to design or build a model for the physical system. As described previously, the lab activities in semester 1 deliberately centered around model testing, and so the models were given but without justification for why or the extent to which the model was appropriate for the physical situation. Semester 2, centered around model building, had statements throughout the semester that were coded as design or build model, but the agency available to groups substantially increased (as evidenced by lighter shades of red in Fig. 4).

For the measurement category, there was similar agency for design or build semesters 1 and 2. However, there were more activities in semester 2 with statements coded as justify design or build, suggesting less available agency. This suggests that the lab instructions shifted toward targeting justification for a design, even though both semesters focused on developing an experimental design.

From semester 1 to semester 2, there was a substantial increase in the available agency for all skills in the comparison category. This is especially apparent in Fig. 6 where the frequency of statements coded in these categories is drastically reduced in semester 2. In semester 1, methods for comparing were provided and justified in statements for one of the lab activities.

The amount of available agency for the refine category also increased from semester 1 to semester 2, though there was generally a lot of agency in this category both semesters. The increase in available agency in semester 2 is only through the removal of coded statements, though there was less available agency (more cuing) for designing refined physical models. This category will be further discussed in the following section.

C. Are lab instructions designed to teach critical thinking through deliberate practice?

With the overall lack of agency and substantial holes in targeted critical thinking skills, the previous analysis suggests that semester 0 did not align with teaching critical thinking skills through deliberate practice. There is also little to no fading of the structure for the skills.

Certainly, the lab instructions in semesters 1 and 2 both had statements targeting repeated practice of critical thinking skills, and, from semester 1 to 2, the available agency

greatly increased. However, according to our scheme and description of deliberate practice in case IV, the lab instructions were not designed to engage students in deliberate practice in either semester. Most obvious is the lack of consistency in fading the number of instructional statements targeting critical thinking skills and in increasing the available agency for those skills.

As indicated previously, agency in the physical category substantially increased from semester 1 to 2, suggesting there were more opportunities to practice these critical thinking skills, but the practice does not seem to be *deliberate* practice in either semester. Statements that were coded as design or build remained with similar levels of agency across lab activities each semester. Statements coded as justify design or build and predictions appeared few times and not in the order predicted for deliberate practice.

For the measurement category, semester 2 appears aligned with deliberate practice in the design or build skill when only examining the level of least agency. However, the number of coded statements fluctuates, making this interpretation unclear. Both the number of coded statements and available agency fluctuates for this skill in semester 1. Semester 2 also more frequently had statements coded as design or build *and* justify design or build in the instructions for the same activity, which may be a better indication of targeted practice designing measurements. However, semester 2 never had statements coded as evaluate data, which makes it unclear whether groups had full or no agency for that skill (depending on whether groups were aware that those decisions were available to them).

From semester 1 to 2, we conjecture that semester 1 was overall better aligned with case IV describing deliberate practice than semester 2 for the comparison category. The second, third, and fourth lab activities of semester 1 included statements coded as all three skills and with relatively little available agency (i.e., many statements and cued with example and none for agency levels). In the fourth activity, all skills were coded but with fewer statements and increased levels of agency, and then, for the final two activities, there were fewer codes and increased available agency. Most of these statements were removed in semester 2, leaving repeated practice over a few activities, but with similar amounts of available agency and without fading over time.

As available agency in the physical category was increased from semester 1 to 2, the available agency in design refined physical model of the refine category was reduced. Likely this reduced agency to refine models was in support of the increased agency to design models in semester 2. This pairing also emerges in semester 1, where statements coded as design refined physical model only appear when there was available agency to groups in design or build in the physical category. However, semester 2 ended with less agency available to groups in the final lab activity. Therefore, it is unclear whether either semester supported deliberate practice.

The number of activities with statements that target refining measurements was reduced from semester 1 to semester 2. In semester 1, the design and justify refined measurement codes somewhat align with the prediction for deliberate practice, with fewer statements coded and with more agency as the semester progressed, unlike semester 2. In semester 2, agency in the refine measurement codes was reduced at the end of the semester. Similar to the pairing for designing and refining physical models, the reduced agency for refining measurements corresponds to an increase in agency for designing measurements.

By looking only at the collapsed codes (activity by activity), the cases outlined in Sec. IID may not capture structures that support deliberate practice within the lab instructions for an individual activity (Figs. 5 and 6). For example, in the electrostatics activity (both semesters), there are two left to right diagonal structures in the order of the codes that suggests two iterations of practicing the full range of critical thinking skills. The first iteration contained more coded statements while the second contained fewer coded statements, but with the same agency. Within this single activity, the lab instructions may have provided more opportunities for students to make decisions on the second iteration, suggesting deliberate practice. There are similar structures in semester 1 for Ohm's law and Faraday's law and in semester 2 for Electricity! Magnetism! Fun! and "what does this thing do?" These results suggest that the scheme and described cases did not fully capture all curricular decisions. Deliberate practice may be integrated at several grain sizes, including within a single activity, across a semester of activities, and throughout a sequence of semesters.

V. DISCUSSION

In this paper, we developed and applied a scheme designed to evaluate lab instructions that aim to teach critical thinking skills through deliberate practice in the context of experimental physics. The scheme draws upon critical thinking, agency, and deliberate practice theories. We applied the scheme to three implementations of electricity and magnetism lab instructions. One of the implementations (semester 0) did not intend to teach critical thinking skills and our application of the scheme to these lab instructions aligned with this intent by demonstrating that student groups had little to no agency available to deliberately practice critical thinking skills (case II). For the two implementations intended to teach critical thinking skills, we found that our curricular design did not fully align with our expectations for teaching critical thinking skills through deliberate practice. A deliberate shift in instructional goals from physical model testing to model building between semesters 1 and 2 appeared to result in lab instructions with improved alignment with the scheme for the physical category but less alignment for the comparison category.

In designing lab instructions for semester 1, we were actively focusing on physical model testing. This design decision shows in the frequency of statements coded in the physical category with little to no available agency for model building. However, from observations and other data sources, we found that students were often trying to confirm a physical model rather than testing limitations, hinting that limited agency around model building also reduced agency available for other categories of critical thinking skills. Previous work has similarly found that students' decision making is reduced in labs that aim to verify a physical model rather than evaluate possible competing models [70]. With a goal to confirm a model, the decisions available to groups may be reduced in the design of measurements and comparisons of measurements to models. Additionally, if a group's result confirms a physical model, then the goal is achieved, leaving no opportunities for refining physical models or measurements [87].

To address this limited available agency, we deliberately designed semester 2 to center around building physical models from experimental evidence. In the coding, design or build of a physical model was still always coded but agency was substantially increased compared to semester 1. However, statements about justifying a physical model remained infrequent despite being an essential component of model building.

Generally, we intended for the lab instructions to cue for justifications but found that explicit cues across all categories were infrequent. Developing the scheme from theory rather than emergent from the instructions and coding only for explicit statements may mean that many characteristics of the instructions were not captured. For example, in Table III, "propose and evaluate several models that may mathematically..." was coded as design or build in the physical category, though may implicitly cue students to justify a model. Another statement in Table II says "If one (or more!) of your models are supported by evidence, design and conduct additional tests that will probe the limits of the model. For example, can you extend the model by testing new variables (e.g., different clothes or rods)?" This statement was coded as design measurement in the refine category but also suggests to a group (but does not cue explicitly) that additional tests and improvements to the physical model require justification. Future work should aim to disentangle how explicit or implicit cuing supports students in justifying their decisions.

Another surprising outcome was the sporadic cuing of specific critical thinking skills. Deliberate practice theory, however, does not directly inform how and when critical thinking skills should be faded or how they coordinate with other critical thinking skills. Our analysis looked at patterns for individual skills, but there appeared to be connections between different skills. That is, as more agency was given for designing or building physical models and measurements, there was less agency for refining them.

It is also unclear how deliberate practice with a single critical thinking skill interacts with the complexity of the specific lab context. For example, agency may be slowly increased while students investigate within a single context but then reduced when students shift to a new context. Our results suggest that this structure may have existed within some lab activities and that our descriptions of cases were at a grain size too large (across activities rather than within the activities) to discern this structure. Three of the six lab activities in semester 1 and three of the five activities in semester 2 include internal structure that resembles deliberate practice. Two of the activities in semester 1 that did not show this internal structure were ultimately dropped in semester 2, and the other activity (Earth’s magnetic field or magnetic fields and coils) was significantly modified. In semester 2, the circuits activity does not show this internal structure, but was developed as review of the prerequisite course material and to familiarize students with relevant equipment. The magnetic fields and coils activity, although modified from semester 1, does not show the internal structure. Our curricular modifications, therefore, may have been attending more to deliberate practice than Fig. 4 suggests.

Figure 7 provides an alternative conjecture to Fig. 2 at the week level (rather than activity level). Figure 7 aligns with deliberate practice often described for helping athletes develop. For example, when training a cross country skier, a coach and athlete may work together to develop particular skills. They may identify a technique change, such as initiating the poling with more power. Then, a specific drill may be developed to isolate particular muscles so that the athlete feels how to make that change, such as raising hands higher in the initial poling phases and bringing the poles down rapidly. The athlete then integrates this skill into endurance and interval training, with reminders and

feedback from a coach such as saying “high hands.” The frequency of the drill is faded as the athlete begins to perform the skill in training and competition without cuing. Under this model, structure would be faded as in a damped oscillator: There is high structure in the first week of the first lab unit, significantly less structure in the second week, then increased again in the first week of the second lab unit (but less than the first lab), and then reduced again in the second session, and so on. We are actively using the scheme to inform changes to lab instructions to align them with Fig. 7.

A. Limitations and open research questions

We analyzed only the written instructions available to students, though the instruction and messaging that students received during lab came from a variety of sources. Curriculum developers control the lab instructions that students receive, but often there is less control over the implementation. There may exist great variation in how lab instructors encourage students to read the instructions, follow alternative directions, and provide information to students [84,85]. Instructional messaging during the lab may greatly affect how the students engage in the activity, so the extent to which written lab instructions influence students’ or instructors’ decision making is unclear.

Furthermore, effective deliberate practice requires a teacher (or coach) to assign task and provide feedback that are tailored to the learner [65,66]. The lab instructions we analyzed were common to all students. Analyzing the in-class support and feedback from the instructor, each student’s group mates, and the experimental apparatus and outcomes are likely necessary to fully understand the role of deliberate practice for learning critical thinking skills in a lab context.

This study did not aim to measure students’ engagement with and development of critical thinking skills, but a crucial piece of curriculum development is comparing student outcomes to instructional intent. Future work will aim to compare students’ use of critical thinking skills during lab activities to the instructional intent. Do students practice the skills that are implicitly or explicitly cued in the lab instructions? How often do students spontaneously use critical thinking skills that are not cued by the lab instructions?

We assumed that the relative difficulties of the lab activities were similar. However, the lab activities included differences in the experimental apparatus, research questions, and statistical tools—presumably, cognitive load (i.e., the number of things to which a person pays attention [88]) varied between activities. Students may have encountered greater difficulties in non-critical-thinking skills in a given activity, which may increase cognitive load and, thus, reduce attention to practicing critical thinking skills.

In designing the scheme, we assumed that critical thinking is context dependent and the relevant context is

FIG. 7. Alternative case for effectively teaching critical thinking skills through deliberate practice. Each box represents one half of a two-week lab activity with the first week providing less available agency compared to the second week.

experimental physics [22,39]. However, our analyses focus only on labs integrated within an electricity and magnetism course, the second course in a three-semester sequence all with learning objectives about critical thinking. Possibly, deliberate practice with the critical thinking skills was well integrated in the previous mechanics course, so the structure in the electricity and magnetism course does not align with our conjectures about deliberate practice. Alternatively, mechanics and electricity and magnetism may be separate contexts within experimental physics, and we do not know how easily students transfer their critical thinking skills between them. We expect students may find practice of critical thinking skills more difficult in electricity and magnetism because they have less physical intuition to draw upon than in mechanics. We also

deliberately incorporated many phenomena that are not traditionally introduced in lecture.

Furthermore, experimentation in physics may have other subcontexts that we have not formally distinguished. For example, in semester 1, we developed lab materials with the intent to engage students in model testing but redeveloped semester 2 around model building. Model testing, where students are familiar with the physical phenomena, may employ different critical thinking skills than model building, where students are less familiar or aware of the relevant physical phenomena.

ACKNOWLEDGMENTS

This work is partially funded by the Cornell University College of Arts and Sciences Active Learning Initiative.

-
- [1] N. G. Holmes and C. E. Wieman, Introductory physics labs: We can do better, *Phys. Today* **71**, No. 1, 38 (2018).
 - [2] M. D. Caballero, D. R. Dounas-Frazer, H. J. Lewandowski, and M. R. Stetzer, Labs are necessary, and we need to invest in them, *APS News* **27** (2018), <https://www.aps.org/publications/apsnews/201805/backpage.cfm>.
 - [3] President's Council of Advisors on Science and Technology, *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics* (United States Office of Science and Technology Policy, Washington DC, 2012).
 - [4] Joint Task Force on Undergraduate Physics Programs, *Phys21: Preparing Physics Students for 21st Century Careers* (American Physical Society and American Association of Physics Teachers, College Park, MD, 2016).
 - [5] AAPT Committee on Laboratories, *AAPT Recommendations for the Undergraduate Physics Laboratory Curriculum* (American Association of Physics Teachers, College Park, MD, 2015).
 - [6] N. G. Holmes, C. E. Wieman, and D. A. Bonn, Teaching critical thinking, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11199 (2015).
 - [7] E. Etkina and A. Van Heuvelen, Investigative science learning environment—A science process approach to learning physics, in *Reviews in Physics Education Research, Volume 1: Research-Based Reform of University Physics*, edited by E. F. Redish and P. J. Cooney (American Association of Physics Teachers, College Park, MD, 2007), https://www.compadre.org/per/per_reviews/volume1.cfm.
 - [8] 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).
 - [9] D. Doucette, R. Clark, and C. Singh, What's happening in traditional and inquiry-based introductory labs? An integrative analysis at a large research university, in *Proceedings of the Physics Education Research Conference, 2018* (American Association of Physics Teachers, Washington, DC, 2018).
 - [10] S. Allie and A. Buffler, A course in tools and procedures for Physics I, *Am. J. Phys.* **66**, 613 (1998).
 - [11] K. Dunnett, M. N. Gorman, and P. A. Bartlett, Assessing first-year undergraduate physics students' laboratory practices: Seeking to encourage research behaviours, *Eur. J. Phys.* **40**, 015702 (2019).
 - [12] Z. Kanari and R. Millar, Reasoning from data: How students collect and interpret data in science investigations, *J. Res. Sci. Teach.* **41**, 748 (2004).
 - [13] S. Allie, A. Buffler, B. Campbell, F. Lubben, D. Evangelinos, D. Psillos, and O. Valassiades, Teaching measurement in the introductory physics laboratory, *Phys. Teach.* **41**, 394 (2003).
 - [14] D. Evangelinos, D. Psillos, and O. Valassiades, An investigation of teaching and learning about measurement data and their treatment in the introductory physics laboratory, in *Teaching and Learning in the Science Laboratory*, edited by D. Psillos and H. Niedderer (Springer, Amsterdam, 2002), Vol. 16, pp. 179–190.
 - [15] M. Séré, R. Journeaux, and C. Larcher, Learning the statistical analysis of measurement errors, *Int. J. Sci. Educ.* **15**, 427 (1993).
 - [16] A. Susac, A. Bubic, P. Martinjak, M. Planinic, and M. Palmovic, Graphical representations of data improve student understanding of measurement and uncertainty: An eye-tracking study, *Phys. Rev. Phys. Educ. Res.* **13**, 020125 (2017).
 - [17] We use “skills” interchangeably with “abilities”; however, other work has provided greater context to the use of these terms, such as Ref. [18].
 - [18] G. Smith, Are there domain-specific thinking skills?, *J. Philos. Educ.* **36**, 207 (2002).
 - [19] R. H. Ennis, A concept of critical thinking, *Harv. Educ. Rev.* **32**, 81 (1962).

- [20] R. H. Ennis, Critical thinking and the curriculum, *Phi Kappa Phi J.* **65**, 28 (1985).
- [21] R. Ennis, Critical thinking: Reflection and perspective Part I, *Inquiry* **26**, 4 (2011).
- [22] M. Lipman, Critical thinking—What can it be?, *Educ. Leader.* **46**, 38 (1988).
- [23] P. A. Facione, The Delphi Report, Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction, in *The Complete American Philosophical Association Delphi Research Report* (The California Academic Press, Millbrae, CA, 1990).
- [24] J. G. Kurfiss, *Critical Thinking: Theory, Research, Practice, and Possibilities* (The George Washington University, Washington DC, 1988).
- [25] R. W. Paul, *Critical Thinking: What Every Person Needs to Survive in a Rapidly Changing World* (Sonoma State University, Rohnert Park, CA, 1990).
- [26] R. Ennis, Critical thinking: A streamlined conception, *Teach. Philos.* **14**, 5 (1991).
- [27] R. H. Ennis, Critical thinking dispositions: Their nature and accessibility, *Informal logic* **18**, 165 (1996).
- [28] B. K. Beyer, Improving thinking skills: Practical approaches, *Phi Delta Kappan* **65**, 556 (1984), <https://eric.ed.gov/?q=thinking+skills&ffl=autBeyer%2c+Barry+K.&id=EJ296775>.
- [29] K. Durkin, The middle way: East Asian master's students' perceptions of critical argumentation in U.K. universities, *J. Stud. Int. Educ.* **12**, 38 (2008).
- [30] D. F. Halpern, Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring, *Am. Psychol.* **53**, 449 (1998).
- [31] M. N. Browne, Coping with future shock: Two styles of critical thinking, *Contemp. Educ.* **49**, 207 (1978), <https://eric.ed.gov/?id=EJ193400>.
- [32] R. D. Dick, An empirical taxonomy of critical thinking, *J. Instruct. Psychol.* **18**, 79 (1991), <https://psycnet.apa.org/record/1991-29405-001>.
- [33] G. L. Hanley, Teaching critical thinking: Focusing on metacognitive skills and problem solving, *Teach. Psychol.* **22**, 68 (1995).
- [34] S. D. Ivie, Metaphor: A model for teaching critical thinking, *Contemp. Educ.* **72**, 18 (2001).
- [35] S. P. Norris, Testing for the disposition to think critically, *Informal Logic* **14**, 158 (1992).
- [36] D. N. Perkins, E. Jay, and S. Tishman, Beyond abilities: A dispositional theory of thinking, *Merrill-Palmer Q.* **39**, 1 (1993).
- [37] P. A. Facione, C. A. Sanchez, N. C. Facione, and J. Gainen, The disposition toward critical thinking, *J. Gen. Educ.* **44**, 1 (1995), https://www.insightassessment.com/var/ezflow_site/storage/pdf/Disposition_to_CT_1995_JGE.pdf.
- [38] D. T. Tiruneh, A. Verburch, and J. Elen, Effectiveness of critical thinking instruction in higher education: A systematic review of intervention studies, *High. Educ. Stud.* **4**, 1 (2014).
- [39] D. T. Willingham, Critical thinking: Why is it so hard to teach?, *Arts Educ. Policy Rev.* **109**, 21 (2008).
- [40] R. H. Ennis, Critical thinking and subject specificity: Clarification and needed research, *Educ. Res.* **18**, 4 (1989).
- [41] 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).
- [42] A. Hofstein and V. N. Lunetta, The laboratory in science education: Foundations for the twenty-first century, *Sci. Educ.* **88**, 28 (2004).
- [43] A. Hofstein and V. N. Lunetta, The role of the laboratory in science teaching: Neglected aspects of research, *Rev. Educ. Res.* **52**, 201 (1982).
- [44] A. Vorholzer and C. von Aufschnaiter, Guidance in inquiry-based instruction—An attempt to disentangle a manifold construct, *Int. J. Sci. Educ.* **41**, 1562 (2019).
- [45] NGSS Lead States, *Next Generation Science Standards: For States, by States* (National Academies Press, Washington, DC, 2013), <https://doi.org/10.17226/18290>.
- [46] M. Ford, “Grasp of Practice” as a reasoning resource for inquiry and nature of science understanding, *Sci. Educ.* **17**, 147 (2008).
- [47] M. J. Ford, Educational implications of choosing “practice” to describe science in the Next Generation Science Standards, *Sci. Educ.* **99**, 1041 (2015).
- [48] U. Zoller and T. L. Nahum, From teaching to KNOW to learning to THINK in science education, in *Second International Handbook of Science Education* (Springer, Dordrecht, 2012), pp. 209–229.
- [49] E. Etkina, A. Van Heuvelen, S. White-Brahmia, D. T. Brookes, M. Gentile, S. Murthy, D. Rosengrant, and A. Warren, Scientific abilities and their assessment, *Phys. Rev. ST Phys. Educ. Res.* **2**, 020103 (2006).
- [50] C. Zimmerman, The development of scientific thinking skills in elementary and middle school, *Dev. Rev.* **27**, 172 (2007).
- [51] J. Osborne, Teaching critical thinking? New directions in science education, *Sch. Sci. Rev.* **95**, 53 (2014).
- [52] A. Bandura, Human agency in social cognitive theory, *Am. Psychol.* **44**, 1175 (1989).
- [53] M. Luck and M. d’Inverno, A formal framework for agency and autonomy, in *Proceedings of the First International Conference on Multiagent Systems* (The AAAI Press, Menlo Park, California, 1995), pp. 254–260, <https://www.aaai.org/Papers/ICMAS/1995/ICMAS95-034.pdf>.
- [54] S. Rodl, Education and autonomy, *J. Phil. Educ.* **50**, 84 (2016).
- [55] M. Emirbayer and A. Mische, What is agency?, *Am. J. Sociology* **103**, 962 (1998).
- [56] R. Paternoster and G. Pogarsky, Rational choice, agency and thoughtfully reflective decision making: The short and long-term consequences of making good choices, *J. Quant. Criminol.* **25**, 103 (2009).
- [57] M. F. Crane, The differential impact of agency and pathway thinking on goal pursuit and university exam performance, *Pers. Individ. Differ.* **58**, 20 (2013).
- [58] A. Bandura, Self-efficacy mechanism in human agency, *Am. Psychol.* **37**, 122 (1982).
- [59] A. Bandura, Exercise of human agency through collective efficacy, *Curr. Dir. Psychol. Sci.* **9**, 75 (2000).
- [60] P. C. Abrami, R. M. Bernard, E. Borokhovski, A. Wade, M. Surkes, R. Tamim, and D. Zhang, Instructional interventions affecting critical thinking skills and dispositions: A Stage 1 meta-analysis, *Rev. Educ. Res.* **78**, 1102 (2008).

- [61] P. C. Abrami, R. M. Bernard, E. Borokhovski, D. I. Waddington, C. A. Wade, and T. Persson, Strategies for teaching students to think critically: A meta-analysis, *Rev. Educ. Res.* **85**, 275 (2015).
- [62] K. Anders Ericsson, R. Th. Krampe, and C. Tesch-Romer, The role of deliberate practice in the acquisition of expert performance, *Psychol. Rev.* **100**, 363 (1993).
- [63] K. A. Ericsson, The influence of experience and deliberate practice on the development of superior expert performance, in *The Cambridge Handbook of Expertise and Expert Performance*, edited by K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman (Cambridge University Press, New York, 2006), pp. 683–703.
- [64] T. van Gelder, Teaching critical thinking: Some lessons from cognitive science, *Coll. Teach.* **53**, 41 (2005).
- [65] D. L. Schwartz, J. M. Tsang, and K. P. Blair, *The ABCs of How We Learn: 26 Scientifically Proven Approaches, How They Work, and When to Use Them* (W.W. Norton & Company, New York, 2016).
- [66] S. A. Ambrose, M. W. Bridges, M. DiPietro, M. C. Lovett, and M. K. Norman, *How Learning Works: 7 Research-Based Principles for Smart Teaching* (John Wiley & Sons, San Francisco, 2010).
- [67] L. S. Vygotsky, *The Collected Works of L. S. Vygotsky*, edited by R. Rieber and A. Carton (Plenum Press, New York, 1934) (reprinted 1987).
- [68] National Research Council, *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*, edited by J. W. Pellegrino and M. L. Hilton (National Academies Press, Washington, DC, 2012).
- [69] K. D. Gutierrez and A. C. Barton, The possibilities and limits of the structure-agency dialectic in advancing science for all, *J. Res. Sci. Teach.* **52**, 574 (2015).
- [70] 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).
- [71] J. A. Schmidt, J. M. Rosenberg, and P. N. Beymer, A person-in-context approach to student engagement in science: Examining learning activities and choice, *J. Res. Sci. Teach.* **55**, 19 (2017).
- [72] 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://www.jstor.org/stable/42993237>.
- [73] R. Lippmann Kung, Teaching the concepts of measurement: An example of a concept-based laboratory course, *Am. J. Phys.* **73**, 771 (2005).
- [74] A. Buffler, S. Allie, and F. Lubben, Teaching measurement and uncertainty the GUM way, *Phys. Teach.* **46**, 539 (2008).
- [75] P. Parreira and E. Yao, Experimental design laboratories in introductory physics courses: Enhancing cognitive tasks and deep conceptual learning, *Phys. Educ.* **53**, 055012 (2018).
- [76] P. Bohacek, M. Vonk, J. Dill, and E. Boehm, Letting students discover the power, and the limits, of simple models: Coulomb's law, *Phys. Teach.* **55**, 380 (2017).
- [77] J. Spears and D. Zollman, The influence of structured versus unstructured laboratory on students' understanding the process of science, *J. Res. Sci. Teach.* **14**, 33 (1977).
- [78] 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).
- [79] S. Pillay, A. Buffler, F. Lubben, and S. Allie, Effectiveness of a GUM-compliant course for teaching measurement in the introductory physics laboratory, *Eur. J. Phys.* **29**, 647 (2008).
- [80] E. Etkina, A. Karelina, and M. Ruibal-Villasenor, How long does it take? A study of student acquisition of scientific abilities, *Phys. Rev. ST Phys. Educ. Res.* **4**, 020108 (2008).
- [81] J. L. Docktor, J. Dornfeld, E. Frodermann, K. Heller, L. Hsu, K. A. Jackson, A. Mason, Q. X. Ryan, and J. Yang, Assessing student written problem solutions: A problem-solving rubric with application to introductory physics, *Phys. Rev. Phys. Educ. Res.* **12**, 010130 (2016).
- [82] N. G. Holmes and E. M. Smith, Operationalizing the AAPT learning goals for the lab, *Phys. Teach.* **57**, 296 (2019).
- [83] M. T. H. Chi, Quantifying qualitative analyses of verbal data: A practical guide, *J. Learn. Sci.* **6**, 271 (1997).
- [84] M. R. Blanchard, S. A. Southerland, J. W. Osborne, V. D. Sampson, L. A. Annetta, and E. M. Granger, Is inquiry possible in light of accountability?: A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction, *Sci. Educ.* **94**, 577 (2010).
- [85] M. Dobber, R. Zwart, M. Tanis, and B. van Oers, Literature review: The role of the teacher in inquiry-based education, *Educ. Res. Rev.* **22**, 194 (2017).
- [86] S. Titscher, M. Meyer, R. Wodak, and E. Vetter, *Methods of Text and Discourse Analysis: In Search of Meaning* (SAGE Publications, London, UK, 2000).
- [87] D. Katchevich, A. Hofstein, and R. Mamlok-Naaman, Argumentation in the chemistry laboratory: Inquiry and confirmatory experiments, *Res. Sci. Educ.* **43**, 317 (2013).
- [88] J. Sweller, Cognitive load during problem solving: Effects on learning, *Cogn. Sci.* **12**, 257 (1988).