Applying a mathematical sense-making framework to student work and its potential for curriculum design

Julian D. Gifford [*] and Noah D. Finkelstein

Department of Physics, University of Colorado Boulder, Box 390, Boulder, Colorado 80309, USA

(Received 26 October 2020; accepted 1 April 2021; published 24 May 2021)

This paper extends prior work establishing an operationalized framework of mathematical sense making (MSM) in physics. The framework differentiates between the object being understood (either physical or mathematical) and various tools (physical or mathematical) used to mediate the sense-making process. This results in four modes of MSM that can be coordinated and linked in various ways. Here, the framework is applied to novel modalities of student written work (both short answer and multiple choice). In detailed studies of student reasoning about the photoelectric effect, we associate these MSM modes with particular multiple choice answers, and substantiate this association by linking both the MSM modes and multiple choice answers with finer-grained reasoning elements that students use in solving a specific problem. Through the multiple associations between MSM mode, distributions of reasoning elements, and multiple-choice answers, we confirm the applicability of this framework to analyzing these sparser modalities of student work and its utility for analyzing larger-scale ($N > 100$) datasets. The association between individual reasoning elements and both MSM modes and MC answers suggest that it is possible to cue particular modes of student reasoning and answer selection. Such findings suggest potential for this framework to be applicable to the analysis and design of curriculum.

DOI: 10.1103/PhysRevPhysEducRes.17.010138

I. INTRODUCTION

Mathematical reasoning and mathematical problem solving play fundamental roles in scientific reasoning. Recently, much attention has been given to the concept of mathematical sense making (MSM) [1–7]. Though there exist multiple and complementary definitions for MSM, most consider it to explicitly involve the simultaneous use of physical and mathematical reasoning. The conversation has begun to progress from defining and describing MSM to assessing engagement in MSM [6]. In prior work [8], we have contributed to the definition and description of MSM, developing a categorical framework that draws on sociocultural theories of cognition [9–11] and the broader definition of sense making synthesized by Odden and Russ [12]. This framework, which we summarize briefly in Sec. I A, provides an operational approach to describing student sense making and is consistent with many other definitions of MSM [1–3,5] and mathematical problem solving more generally [13–21].

[*]Julian.Gifford@colorado.edu

Previously, we have applied this framework to describe the nuanced reasoning exhibited by students in extended, focus group (interview) settings [8,22]. In the present work, we expand this analysis by applying the framework to new modalities of student work, in particular the sparser data streams of individual written responses and multiple-choice answers. We note that this is not only a different data stream than before (e.g., different content) but a different *kind* of data. In shifting to these new modalities, we also consider a much larger sample size ($N \sim 100$). The expanded analysis corroborates our prior work and is a necessary step to establish the utility of the MSM framework as a tool for describing student reasoning across multiple modalities of student work and across sample sizes. In particular, we establish that the framework can be applied productively to data streams more commonly observed by physics educators.

Recently, the conversation in the literature has extended beyond descriptions of MSM and has begun to assess student engagement in it. The present work establishes that the framework can contribute to these discussions and could be a useful tool for researchers and educators in describing and assessing reasoning across modalities of student work. We also begin to establish a theoretical foundation for the framework to be used in the design and assessment of curriculum. Applications to curricular design are an important practical and theoretical step as the discussion in the physics education research (PER) and science education communities moves beyond describing and assessing MSM and begins to develop ways to support student engagement in MSM.

As an attempt to understand some of the mechanisms underpinning these reasoning modes, and as a first step towards the predictive use of this framework, we also look at the association between MSM modes and the use of finer-grained cognitive tools that we term *reasoning elements*. The varied, contextually specific use of these reasoning elements provides a more detailed description of widespread student reasoning around a given topic and brings greater specificity to the associated MSM modes.

These reasoning elements are consistent with a *knowledge-in-pieces* [23] or resources perspective [24,25] and are akin to *facets* [26,27]. The reasoning elements, either individually or specific clusters, can be associated with both the MSM modes and multiple-choice answers leading to bidirectional mappings between reasoning elements, MSM modes, and multiple-choice answers. These associations suggest that student sense making (i.e., engagement in the various MSM modes) and answer selection can be scaffolded by targeted activation of particular reasoning elements. We test this predictive hypothesis here, the results of which offer preliminary validation of this approach and support the potential for this framework to be applied to the analysis and design of curricula and other instructional interventions.

The present work is conducted in the context of the photoelectric effect and much attention is given to the development of, and student reasoning on, a novel instructional item we call the "KE vs λ task." This task involves multiple representations (symbolic and graphical) and the interplay between mathematical and physical interpretations of mathematical formalisms. Student reasoning on this task is an interesting result in and of itself, but as this is not the primary focus of the paper a large part of this discussion is placed in the Appendices. However, much of the analysis presented in the main text does focus on the task itself and associated reasoning. Grounding our discussion in the items of the task allows us to establish connections between relevant reasoning and the MSM modes. Our goal here is not to make claims about student sense making in general, but to show that the framework can be used to validly associate sense-making modes and answer choices on a given problem.

Over the course of this study we (i) demonstrate that the MSM framework can be applied to analyze student written work and to make inferences from multiple choice responses allowing for the analysis of larger sample sizes ($N > 100$), and (ii) establish associations between MSM modes, reasoning elements, and answers. In addition to these two points, which attend to the descriptive utility of the framework, we also begin to attend to the potential predictive utility of the framework. To that end, we show that (iiia) the targeted cueing of specific reasoning elements has a predictable role on both student answers and associated MSM modes, and (iiib) that this cueing likely depends on the representational format of the prompts provided. These results provide preliminary indication that this framework can be applied to analyze and design curricula, which will be the subject of future work.

A. Framework of MSM

Here, we briefly summarize the categorical MSM framework. For a more in-depth discussion of scientific sense making, mathematical sense making, and the categorical framework, see our prior work [8,22] and Russ and Odden's review of scientific sense making [12].

Odden and Russ offer a general definition of sense making as the "process of building or revising an explanation in order to 'figure something out'—to resolve a gap or inconsistency in one's understanding." This general definition is made more concrete as they show that sense making has been conceptualized in the science education literature in three distinct but complementary ways: as (i) an epistemological stance towards science learning, (ii) a cognitive process, and (iii) a specific discourse practice. Our categorical framework for *mathematical* sense making (in physics) builds from the cognitive process strand of scientific sense making and draws on Vygotsky's notion of mediated cognition [9].

In mediated cognition, a mediator (some sort of tool) is used to aid the subject's understanding of an object, e.g., using the equation $\vec{F} = m\vec{a}$ (tool) to predict what will happen when you push on a heavy box (object). This mediated understanding is a complement to our direct understanding of the object (in this case not the box alone but the scenario of a box being pushed on by you) which is built up from experience and experimentation. Mediated cognition is often represented as a triangle with the nodes *subject, tool*, and *object*, as shown in Fig. 1, where the base of the triangle represents a direct understanding of the object by the subject and the two legs represent a mediated (by the specific tool) understanding.

The MSM framework defines four basic modes of mathematical sense making in physics, that are

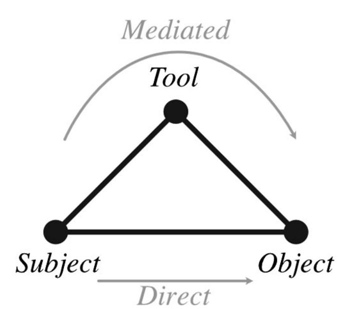

FIG. 1. A representation of mediated cognition, showing the mediated (by a tool) and direct pathways by which a subject interacts with an object. Drawn from Vygotsky [9].

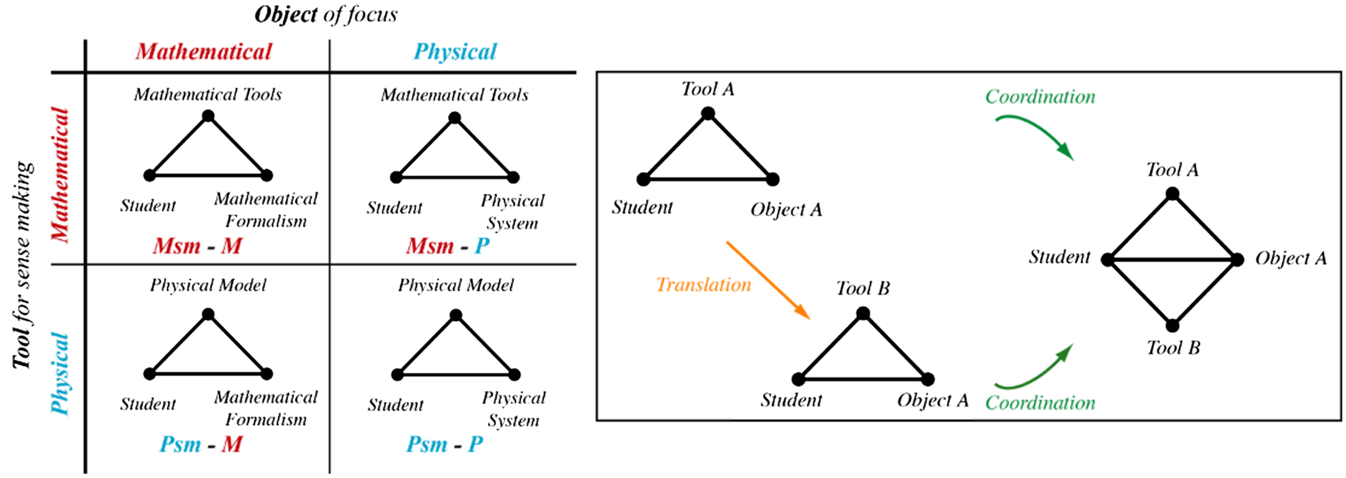

FIG. 2. Left: the four modes of the MSM framework, defined by the *tool* and the *object* of sense making. Right: a sense-making diagram, showing the processes of *translation* and *coordination* that lead to a coordinated reasoning structure.

characterized by the *object* of sense making (whether it is mathematical or physical) and the *tool* used to mediate that sense making (mathematical or physical). The four modes of the framework are shown in the left of Fig. 2. For a more detailed explanation of mediated cognition and the MSM modes see our prior work [8,22], here we present only a brief example.

In an Msm-P mode of reasoning (*mathematical* sense making of a *physical* object)—perhaps the canonical mode of MSM in physics—mathematical formalisms are used as a tool to understand physical phenomena. Consider an electron in a finite double square well, a sketch of this potential and the first two energy eigenstates are shown in Fig. 3. The Schrödinger equation

$$\frac{-\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi(x) = [E - V(x)]\psi(x)$$

can be solved to give the wave function [$\psi(x)$], and this function can be leveraged to understand the behavior of the electron; for example, where it is most likely to be found, how excited a state the electron is in, or the probability that, if released in one well, it will tunnel to the other. While this

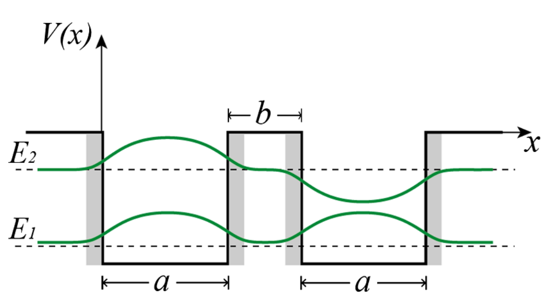

FIG. 3. A representation of the finite, double square well, and the first two energy eigenstates (wave functions).

abstract mathematical formalism can be used to understand the behavior of the electron, physical reasoning can also be used to understand *why* the wave function has the shape that it does, this would be a Psm-M mode of reasoning (*physical* sense making of a *mathematical* object). In this case, one could consider that $E - V(x)$ is the *kinetic energy*, which has a negative value in the tunneling region. This negative kinetic energy is not classically possible and implies that the electron does not have the energy to be in this location, which means that the wave function should be exponentially decaying rather than oscillating.

While it is possible that entire episodes of student reasoning can be categorized into one of these four modes, our prior work showed that often sense making, the process of figuring something out, involves building connections between these modes. If the modes are seen as "atoms of reasoning," then the larger process of sense making leads to a reasoning structure that can be considered a "molecule of sense making." In prior work [8] we identified three prevalent processes by which these larger scale structures of sense making are constructed: *translation*, a shift between any one of the four modes; *chaining*, where the object of sense making is used subsequently as the tool for future sense making; and *coordination*, the combination of two different tools (MSM modes) to understand the same object.

The generation of this larger sense-making "molecule" can be represented visually in a *sense-making diagram*. A sense-making diagram is shown on the right of Fig. 2 and shows the generation of a coordinated reasoning structure. While most definitions of mathematical sense making in the context of physics involve leveraging multiple cognitive and representational tools (both mathematical and physical) to develop an understanding of physical phenomenon, our framework makes this explicit with the process of coordination. One particular example of coordinated reasoning occurs in the context of the photoelectric effect, where physical reasoning imposes constraints on the use of a mathematical tool.

B. Photoelectric effect background

The present study occurs in the context of the photoelectric effect. Here, we present a detailed examination of the physical and mathematical principles that the student will be drawing on in answering the KE vs λ task, which is discussed in Sec. II B and Appendix A. We also specifically highlight how the coordination of physical and mathematical reasoning is required to properly apply the relevant mathematical formalism to make sensible predictions of the physical system.

In the canonical photoelectric effect experiment, light of frequency f is incident on a metal plate of work function Φ, where the work function is a measure of how energetically bound electrons are to the metal. A schematic showing the simplified experimental setup is shown in Fig. 4. Regardless of the intensity (brightness) of the light, electrons are not ejected from the plate if the frequency is below a threshold ("cutoff") frequency. Above this frequency, the kinetic energy of the ejected electrons is again independent of the intensity, instead depending directly on the frequency. The common expression for the max kinetic energy (KE) of an ejected electron is given by [28]

$$KE_{\mathrm{max}} = hf - \Phi. \tag{1}$$

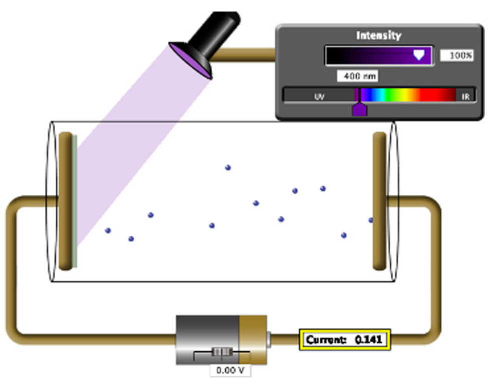

FIG. 4. A schematic of the photoelectric effect experiment, taken from the PhET simulation [29]. Light incident on the metal plate ejects electrons which travel to the right. The rate at which electrons reach the right plate is measured by the ammeter, the max current is a measure of the ejection rate and is independent of the battery voltage when the voltage is positive. If the battery is turned around (the voltage is negative), the electrons are slowed down and the less energetic electrons no longer cross the gap. The voltage at which no electrons make it to the right plate is called the *stopping potential* (V_0) and is a measure of the kinetic energy of the most energetic electrons ($\mathrm{KE}_{\mathrm{max}}$).

However, this is an incomplete description of the phenomenon, and Redish and Kuo note [7] that a mathematician would more likely (and more correctly) include the Heaviside step function $[\theta(x - x_0)]$ to write

$$KE_{\mathrm{max}} = (hf - \Phi)\theta(f - \Phi/h) = \begin{cases} 0, & \text{for } f < \frac{\Phi}{h} \\ hf - \Phi, & \text{for } f \geq \frac{\Phi}{h} \end{cases}, \tag{2}$$

which accurately describes both the linear frequency dependence of the energy and the cutoff frequency, the two most essential aspects of the photoelectric effect in terms of the photon model. This difference is well exemplified by the graphical representations of these two functions, as shown in the left of Fig. 5. The first function (the red dashed plot), suggests that for frequencies below the threshold frequency the kinetic energy of the electrons will be *negative*, while the later (the solid blue plot) correctly shows that the kinetic energy is *zero* below the threshold frequency, as no electrons will be ejected. As the piecewise nature of this function is rarely made mathematically explicit, this is an instance where the simultaneous coordination of physical and mathematical reasoning is required to accurately apply this equation to a physical system.

Because of the hidden mathematical structure and the rich connections between the physical system and the symbolic and graphical representations used, the photoelectric effect is an excellent context to consider mathematical sense making. To investigate student reasoning and the interplay between mathematical and physical sense making, we extended the traditional discussion of the photoelectric effect (which focuses on the *frequency* of light) and considered the maximum kinetic energy as a

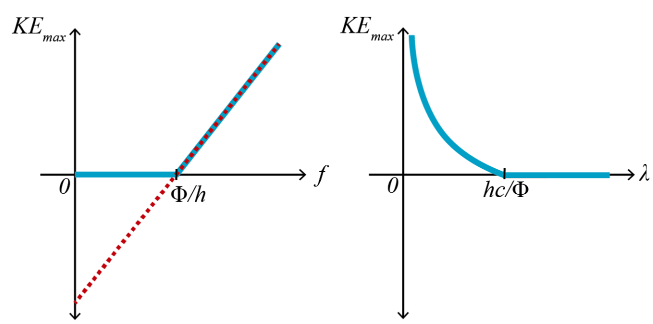

FIG. 5. Left: Plots of kinetic energy versus frequency as defined by Eq. (1) (dashed and plotted in red) and Eq. (2) (solid and plotted in blue, the canonical representation), which differ in their implicit [Eq. (1)] and explicit [Eq. (2)] inclusion of the cutoff frequency. Right: A plot of the max kinetic energy of the ejected electrons versus the *wavelength* of light as described by Eq. (3). Note that this is a noncanonical equivalent representation to KE vs f and that the minimum cutoff frequency has become a maximum cutoff wavelength.

function of the *wavelength* of light. This is a noncanonical approach, and a plot of KE vs λ does not appear in the textbook (Knight [28]), or any modern-physics textbook that we know of. Furthermore, in a review of the PER literature concerning the photoelectric effect, we found no mention of a *plot* of KE vs λ, and only a single mention of an instructor explicitly including a discussion of the symbolic function $\mathrm{KE}(\lambda)$ [30]. In fact, the bulk of the literature [30–33] has built on work by Steinberg, Oberem, and McDermott that led to the development of the computer program *photoelectric tutor* [31] and focuses on the experimental design and the creation and interpretation of IV (current versus voltage) plots. More recent work has continued to taxonomize the common difficulties (predominantly conceptual) that students (or pre or in-service teachers) face [34,35] and investigated the hidden mathematical grammar of the photoelectric effect [7], but little work has been done to study MSM, and particularly the coordination between mathematical and physical reasoning, in the context of the photoelectric effect.

Formally, the function $\mathrm{KE}(\lambda)$ can be determined by substituting the frequency for the wavelength given that, for an EM wave in vacuum, $c = f\lambda$. Written piecewise to explicitly include the threshold wavelength, this function is

$$\mathrm{KE}_{\mathrm{max}} = \begin{cases} \frac{hc}{\lambda} - \Phi, & \text{for } \lambda \leq \frac{hc}{\Phi} \\ 0, & \text{for } \lambda > \frac{hc}{\Phi} \end{cases}. \tag{3}$$

This substitution must be made for the bounds as well as the function and, due to the inverse relationship between frequency and wavelength, the *minimum* cutoff frequency now corresponds to a *maximum* cutoff wavelength. While this is a somewhat trivial algebraic substitution, it is worth noting that the idea of an inverse is *not* trivial and poses many conceptual issues for students [36]. Additionally, in conversations with students, we realized that this algebraic substitution provides another example of the reasoning associated with what Redish has called "Corinne's shibboleth" [7,37]—which highlights differences between the "dialects" of math-in-mathematics and math-in-physics.

In his analysis of Corinne's shibboleth [7], Redish introduces the idea of a *physical function* as opposed to a *mathematical function*. A physical function, he argues, represents a particular quantity that depends in some way on the specified variables, for example the kinetic energy as a function of frequency [$\mathrm{KE}(f)$] or wavelength [$\mathrm{KE}(\lambda)$]. Where the label KE indicates the kinetic energy, regardless of the functional dependance on the variable. A mathematical function, on the other hand, establishes a particular operation to be performed on arbitrary quantities. Thus, the label KE would represent the specific mathematical operation "multiply the variable by h and subtract Φ" giving $\mathrm{KE}(\lambda) = h\lambda - \Phi$—which no longer accurately describes the kinetic energy. This exact reasoning was evident in some student responses to the KE vs λ task. Further discussion on the KE vs λ task, Corinne's shibboleth, and this "photoelectric shibboleth" can be found in Appendix A.

II. METHODS

This work is part of a larger project aimed at describing, understanding, and promoting student reasoning and mathematical sense making. To continue the investigation of student reasoning regarding the photoelectric effect, and to test the utility of the categorical framework in describing student reasoning based on written work and multiple-choice answers (versus the original interview protocols), we developed the KE vs λ task, and implemented it in a three-stage fashion in two separate semesters.

This is a naturalistic study [38] in that the data come in the form of homework and exam responses collected naturally through the course of the semester. While the prompts were designed to elicit MSM, there was little done to control for external factors and students were not told explicitly that these responses would be collected as data for this particular study. In this sense, these data are akin to observations made on students in a natural educational environment. This study does differ from a standard naturalistic study in comparing aggregate student responses between semesters 1 and 2. Traditionally, there is no manipulation to the environment; however, in comparing between semesters we actively altered the form of the prompt to observe a natural—but anticipated—response.

A. Course context and data collection

The data presented come from two semesters of a middle-division modern physics course at CU Boulder, primarily intended for engineers. In both semesters the second author (N. D. F.) was the instructor of record and the first author (J. D. G.) was a co-instructor and course co-designer, playing a substantial role in both curriculum design and instruction. There were 114 students enrolled during semester 1, and 68 students enrolled during semester 2; in both courses $\sim$17% of students were female. The majority of students are mechanical, electrical, or computer science engineers, and most are upperclassmen.

In each semester, the KE vs λ task was implemented in three stages for all students. First as a multiple-choice pretest survey administered before any instruction (week 1) for which students were given homework credit for completion; second, as a two-part homework question—a repeat of the multiple-choice pretest with a free response follow up—(due week 3) following interactive instruction on the photoelectric effect (including ConcepTests [39] involving the symbolic expression $KE = hc/\lambda - \Phi$ and a tutorial that calls attention to the implicit limitations of the expression $KE = hf - \Phi$); and finally as a question on the first midterm exam (week 5).

The pretest and homework questions were identical over the two semesters ("version A" of the task, see Sec. II B below) and instruction was essentially the same; however, the exam questions were substantially different in semester 1 (version B) and semester 2 (version C). Both the pretest and homework questions were administered online (using Qualtrics and Canvas, respectively) and student responses were collected using the course management system. The exam question in semester 1 was a required free response question and handwritten student work was collected and scanned. In semester 2 the exam question was a required multiple-choice question, and responses were collected via scantron. Statistical hypothesis testing was conducted on these data sets, discussed in greater detail in the analysis sections below. While Pearson chi squared (χ^2) is a common statistic for determining the significance of contingency tables, it loses considerable statistical power for tables containing cells with $N < 5$ [40]. In these cases, we instead (or in addition) use the Fisher exact test [41].

At the end of semester 1, two 3-student focus group interviews were conducted by the first author (J. D. G.). Included in these interviews was a question similar to version B of the KE vs λ task. These data were used to validate the code book of reasoning elements (see Sec. II C 3) and for comparison with the written work analyzed in Sec. III A. Beyond this validation of the sparser data streams (written and multiple-choice responses) presented in this paper, these data will not be discussed in the present work; however, a detailed analysis is available in prior work [8]. No focus groups were conducted during semester 2.

The goals of data collection differed slightly between the two semesters. The focus of semester 1 was to expand and validate the prior use of the MSM framework by triangulating among multiple modalities of student data (focus group interviews, written work, multiple-choice answers). This addresses our first two research questions, which attend to the descriptive utility of the framework. The primary goal here was to determine how and if the categorical framework can be productively applied to describe student reasoning and sense making using data from these sparser data streams, and if the signal seen in these data is consistent with the more fine-grained and richer data obtained from focus group interviews. To that end, the KE vs λ task involves multiple modalities of student work (Multiple-choice and written) and includes iterations framed both more mathematically (pretest and homework) and more physically (exam). These data are used to substantiate this association and, as shown below, we have indication to believe that student multiple-choice answers can be correlated to specific modes of the MSM framework.

Following our preliminary analysis of semester 1, in semester 2 we further corroborate the associations between individual reasoning elements and multiple-choice answers. This further addresses the descriptive utility of the framework (our first two research questions) and also begins to address the predictive utility of the framework (our third research question). Specifically, we test the hypothesis that cueing (the style of prompt) can directly influence the use of specific reasoning elements, and so is associated with answer choice and hence correlated with a particular MSM mode. In line with Podolefsky's work on representation and analogy [42], we designed a multiple-choice exam question presenting the KE vs λ task in a physics context using a graphical representation. This is intended to scaffold student use of two specific reasoning elements while suppressing the use of a third. As will be discussed below, the use of these reasoning elements has a predictable effect on student answers.

B. The KE vs λ task

As noted above, there were three versions of the KE vs λ task. Version A was implemented on the pretest and homework for both semesters, it is framed in a mathematical (rather than physical) context and student responses are multiple choice, with a free response follow up on the homework. Version B was implemented on the exam in semester 1, it is framed in a physical context and student responses are free response. Version C was implemented on the exam in semester 2, it is framed in a physical context and student responses are multiple choice. Each of these tasks will be discussed briefly here to establish context for the analysis of student data. An in-depth examination of these tasks, including both their design and analysis from an MSM perspective, is presented in Appendix A. The bulk of our analysis will be focused on student responses to the exams in semesters 1 and 2, though we draw on the results of the pretest and homework for two reasons: (i) to establish a statistical baseline allowing for a comparison between the two semesters and (ii) to support our discussion of overall framing in scaffolding student engagement in MSM. A detailed discussion of student responses to the pretest and homework is provided in Appendix B.

1. Version A: Pretest and homework

Version A, shown in Fig. 6, asks students to make an algebraic substitution to a given piecewise function and plot the result. The five distractors were designed based on preliminary results from prior focus group studies and several assumptions of possible student difficulties based on this algebraic substitution. Our intention was to establish how students used the mathematical formalism when isolated from the physical context of the photoelectric effect.

In the context of the photoelectric effect there is an unambiguously correct answer: plot (f). However, there are actually two "correct" answers to this question, depending on whether $f(x)$ is treated as a *mathematical function* or a *physical function*. When treated as a mathematical function, neither the function $f(x)$ nor the variables x and y represent

Consider the following function:

$$f(x) = \begin{cases} 0, & for\ x < B/A \\ Ax - B, & for\ x \geq B/A \end{cases}$$

If $x = \frac{D}{y}$ which of the following is a qualitatively accurate plot of $f(y)$?

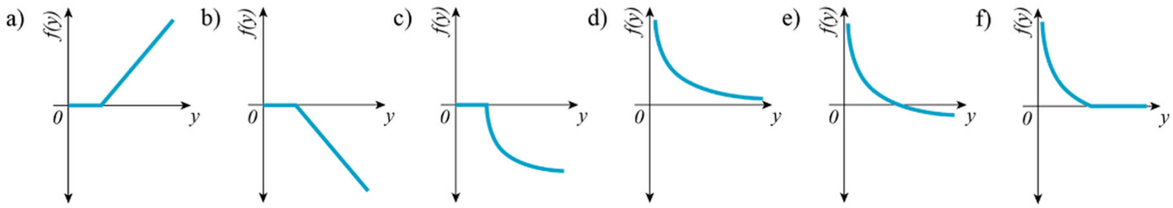

FIG. 6. Version A of the KE vs λ task is framed in a mathematical context and responses are multiple choice. This question was asked in both semester 1 and semester 2 on the pretest (week 1) and again on the second homework (week 3). The homework included a free response follow up asking students to link this abstract mathematical formalism to the photoelectric effect experiment.

specific physical quantities, and so regardless of the variable used the function f specifies the relationship represented by plot (a). The effect of framing will be discussed further below.

Ultimately, version A requires students to select a *graph*—an object that is fundamentally mathematical in nature, though inarguably contains physical significance given appropriate context. All five of the distractors were designed based on assumed difficulties with the mathematical formalism (conflation of inverse and negative linear, failure to switch or include the bounds, etc.). Because of the form of the distractors and the lack of physical context, it is expected that the framing of this question cues students towards an Msm-M mode of reasoning, where the generation (selection) and understanding of the plot is based primarily on the use of mathematical tools and procedures.

Physics follow-up (HW): On the homework, version A of the KE vs λ task was repeated as shown but with a free response follow-up question. The follow-up stated that the mathematical formalism is a relevant description of some aspect of the photoelectric effect experiment and asked students to consider relevant parameters in the experiment (current, voltage, kinetic energy, work function, incident light, etc.) and explain what this formalism represents. This follow-up question provides a physical system and was intended to encourage Psm-M reasoning such that students would draw on relevant knowledge of the photoelectric effect to contextualize the mathematical formalism. The implementation of the pretest and homework questions were identical across the semesters, other than that the ordering of the graphs was changed on the homework in semester 2 to mitigate direct copying of prior solutions.

2. *Version B: Semester 1 exam*

Version B, shown in Fig. 7, presents a physical system—a specific photoelectric effect experiment—and has two parts. Part i provided students with specific values for the frequency of light and the work function of the plate and asked for an explicit calculation of the maximum kinetic energy of the electrons. The values given were such that the frequency was below the threshold frequency and so the correct answer to part i is *zero*, though simply calculating the value from KE $= hf - \Phi$ yields a negative number. Our intention was to scaffold engagement in a Psm-M mode of reasoning, using the physical impossibility of "negative kinetic energy" to interpret an unusual mathematical result in terms of the physical system. The intended outcome was that students would be cued into recalling the cutoff frequency and limitations of the standard symbolic

In a particular photoelectric effect experiment, light of frequency $f = 8 \cdot 10^{14}$ Hz is incident on a metal plate with work function $\Phi = 3.5$ eV.

i) What is the maximum kinetic energy of the ejected electrons in this experiment? Explain your reasoning.

ii) The experiment is repeated with light of many different wavelengths. On the axes below make a plot of the maximum kinetic energy versus the *wavelength* (not frequency) of light used. Indicate any asymptotes, limits, and/or interesting features. Explain why the plot has the shape that it does, and what this tells you about the photoelectric effect experiment.

FIG. 7. Version B of the KE vs λ task is framed in a physical context and student responses are free response including a sketch of the graph. This question was asked on the exam (week 5) in semester 1.

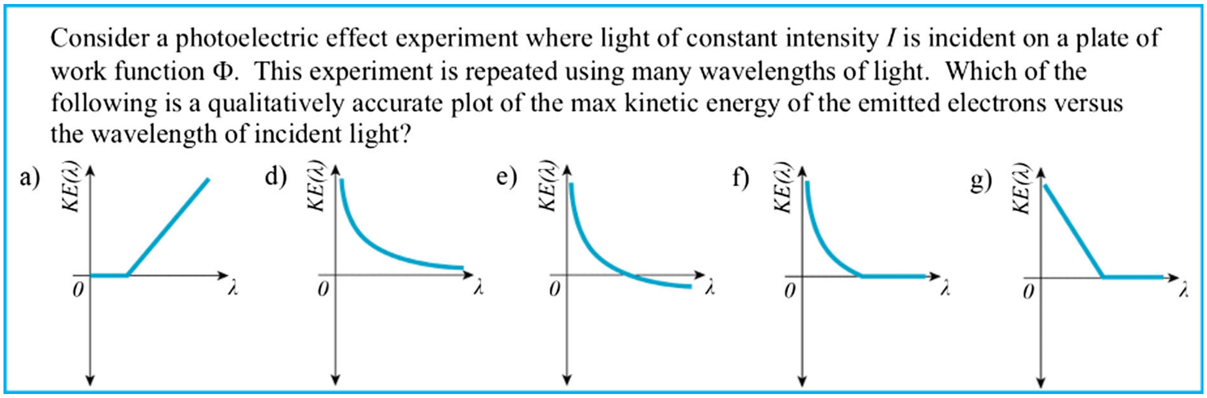

FIG. 8. Version C of the KE vs λ task is framed in a physical context and student responses are multiple choice. This question was asked on the exam (week 5) during semester 2. Note that the labels [(a) and (d)–(g)] are used here to consistently indicate the associated plots throughout this paper; however, they were not labeled as such on the exam.

formalism (equation (1). This approach was less effective than anticipated, as many students simply accepted a negative value for KE. This suggests either direct plug and chug (which is not necessarily considered to be sense making) or engagement in an Msm-M mode of reasoning, where the equation $\mathrm{KE} = hf - \Phi$ is treated as a mathematical equation with no physical significance. The failure of this type of cueing prompted our study of semester 2 using a different kind of cueing.

Part ii asked students to make a plot of $\mathrm{KE}(\lambda)$, indicating any special features and discussing what the plot suggests about the photoelectric effect experiment. Though Part ii of version B is "isomorphic" in content to version A, there are several relevant differences between them. In particular, the format of version B is free response rather than multiple choice, and so the task is to *create* a graph rather than select one. Additionally, the problem statement does not explicitly provide the mathematical form of $\mathrm{KE}(f)$ in either the traditional or piecewise notation. Instead, the physical system is specified, and the canonical symbolic expression $\mathrm{KE}(f)$ is implicitly requested. Because of this framing, the physics context version is, arguably, a more difficult question that less strongly cues a direct Msm-M mode of reasoning. While these two versions are not directly comparable, by the time of the exam (the fifth week of the semester) students had seen version A twice and extensive solutions to the HW detailing both the algebraic substitution and physical implications for the photoelectric effect had been available for two weeks.

3. Version C: Semester 2 exam

To test the hypothesis that a graphical representation presented in a physics context will provide stronger cueing for the cutoff frequency or wavelength, a multiple choice, physics context version was developed (shown in Fig. 8). Version C presents a physical scenario, a repeated photoelectric effect experiment with varied wavelength, and asks for the corresponding plot of KE vs λ. As this is a multiple-choice format the task is to select rather than generate a graph, and so is an arguably easier task than version B. Though the answers were presented to students as choices (a)–(e), for clarity in this paper we maintain a consistent labeling scheme for the plots, and so the labels in Fig. 8 read (a) and (d)–(g). The distractors include plot (a), the more mathematically motivated plots (d) and (e), and a more physically motivated distractor plot (g). As will be discussed below, plot (g) was emergent from student responses to version B on the semester 1 exam.

In discriminating between the answers there are three relevant characteristics: (i) is the plot straight or curvy, (ii) increasing or decreasing, and (iii) is there a cutoff or can it be negative? Our hypothesis is that addressing these questions in the context of the photoelectric effect now more strongly cues students into a Psm-M mode of reasoning than the calculation of part i of version B.

To establish similarity between the student populations, responses to the pretest and homework (both version A) were compared between the two semesters. Based on Fisher's exact tests, we do not reject the null hypothesis that the distribution of responses and semester are independent at the $p = 0.05$ level for the pretest, or even at the $p = 0.1$ level for the homework. As such, it is reasonable to assume that these two populations of students, who had experienced effectively identical instruction, were similar going into the exam and that the only significant difference was the format of the exam question. Further discussion of the student populations and responses to the preteset and homework can be found in Appendix B.

C. Coding student responses

In the analysis below we take a mixed-methods approach, triangulating across two modalities of student work: written responses, and multiple-choice answers. This analysis is compared with an extended analysis of student reasoning from the focus groups, discussed in a prior paper [8,22]. While the prior analysis explored the fine-grained shifts in student reasoning on longer problems, in the present analysis the MSM framework is applied to describe the dominant reasoning modes present in student written work and to associate multiple-choice answers with individual reasoning elements, clusters of reasoning elements, and a corresponding MSM mode. To that end, student responses to the semester 1 exam (version B) were coded in three ways: for the answer (plot drawn), for the MSM mode associated with their written work, and for the reasoning elements present in their response.

Responses to the pretest and homework were collected for all students in both semesters, and are used primarily to establish statistical similarity between the two semesters. Our primary arguments do not draw heavily on these results, but a detailed analysis of these data (for semester 1) that explores the utility of the framework in considering change over time is presented in Appendix B.

1. Coding plots (multiple-choice answer)

For comparison with the pretest and HW, the exam responses were first coded for the plot drawn, using the lettered plots from version A [plots (a)–(f)] and "other" as an initial code book. From the responses categorized as other, initially 30% of all responses, two new codes were defined—plots (g) and (h), shown in the right of Fig. 9. With the addition of plots (g) and (h) only 8% of responses remain coded as other, and many of these plots have a strong resemblance to the lettered plots (a)–(h), albeit with some notable difference. For example, the left side of Fig. 9 shows a response that strongly resembles plot (f) but was coded as other due to the unusual jump to zero and the vertical asymptote at a nonzero value of wavelength.

2. Coding MSM modes

All semester 1 exam responses with a graph coded as plots (d)–(h) (99/114 responses) were coded by the first author for a dominant MSM mode, labeling each explanation as Msm-M, Psm-M, or coordinated—where coordinated implies *both* Msm-M and Psm-M reasoning. Though an additional code of other was meant to capture any responses that could not be coded as such, this code was ultimately not necessary. A subset of these responses (33/96) was also coded by the second author to establish inter-rater reliability on this assessment of MSM mode.

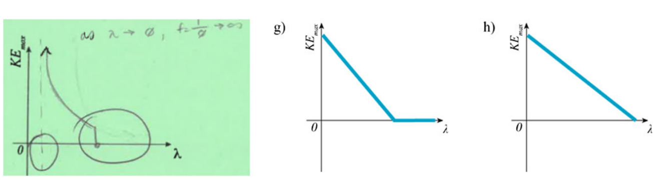

FIG. 9. Left: An example of a plot coded as other that bears a strong resemblance to one of the lettered plots. The two circles are marks made by the grader. Right: Examples of the novel plots (g) and (h), emergent from the version B free response exam data.

Before any discussion, there was 88% agreement (29/33) on the codes, and after a brief discussion the authors reached 100% agreement. As the semester 2 exam (version C) was entirely multiple choice, no MSM coding was conducted on these data.

3. Defining and coding reasoning elements

In addition to the holistic, if coarse-grained, coding of MSM reasoning structure, all written work on the semester 1 exam was also coded for the presence or absence of particular tools or approaches which we have termed "reasoning elements." These reasoning elements are in line with a resources perspective [24,25], which focuses on cognitive elements rather than full blown concepts [23]. They may exist on a scale larger than *p* prims [23], and are akin to *explanatory* primitives [43] or facets [26,27]. Since they are explicitly dependent on the context, they can be fundamentally incorrect and are not necessarily a "given" in the sense of *p* prims or *e* prims. However, if student work included an equation resembling $\mathrm{KE}(\lambda) = hc/\lambda - \Phi$, their response was coded as including the *full equation* reasoning element even if the equation was written or used incorrectly.[1] We anticipate that unique combinations of the various reasoning elements are suggestive of different modes of MSM. That is, the prevalence and/or combination of certain reasoning elements used in solving a particular problem may distinguish between engagement in the different MSM modes.

The development of this code book of reasoning elements was an iterative process. Initially student responses were coded only for the graph drawn, and brief notes were taken on the explanations provided by each student. From these notes a preliminary set of reasoning elements was identified, and each student response was then coded as either including or not including each of these reasoning elements. During this second round of coding three new reasoning elements were identified and a third round of coding was conducted, after a discussion between the authors, using these new elements. Following the third round of coding J. D. G. met with another professor at CU Boulder who was teaching an upper division Quantum Mechanics course. Version B of the KE vs λ task was presented to the professor and many hypothetical student responses were considered. After this period of hypothesizing, J. D. G. presented the eight coded plots (a)–(h) and the code book of reasoning elements. The code book captured the entirety of the hypothetical responses

[1]While this framing is in line with a resources perspective, to avoid conflation with the specifics of the resources literature, we term them reasoning elements rather than resources and note that while the use of a tool is similar to the activation of a resource (or cluster of resources) these cognitive processes are likely not identical.

TABLE I. Code book of reasoning elements, emergent from student responses to version B of the KE vs λ task on the semester 1 exam.

Reasoning element	Description	Example
Full equation—Mathematical	Student work included a full functional expression for KE(λ). This RE was still coded as present even when the mathematical expression was incorrect.	$\mathrm{KE}_{\mathrm{max}} = \frac{hc}{\lambda} - \Phi$
Inverse (s)—Mathematical	Any *symbolic* relation indicating that wavelength is inversely realted to KE *or* frequency. Note: Any response that included *full eq* automatically includes *inverse (s)*.	$\mathrm{KE} \propto 1/\lambda$ OR $f \propto 1/\lambda$
Asymptotes—Mathematical	Any discussion (or calculation) of the asymptotes or limiting values of the plot for $\lambda = 0$ and/or $\lambda \to \infty$. Note: this is distinct from the *cutoff* codes.	"*As λ gets large the kinetic energy approaches zero*" OR "*as the wavelength gets shorter the* KE *blows up*"
Cutoff (c)—Coordinated	A calculated value for the cutoff wavelength above which there is no ejection ($E_{\mathrm{photon}} = \Phi$ OR $\mathrm{KE} = 0$).	$\lambda_{\mathrm{cutoff}} = \frac{hc}{\Phi}$ OR 354 nm
Cutoff (v)—Physical	A verbal discussion that there exists a cutoff wavelength above which electrons will not eject.	"*When λ is too large (above the threshold frequency) no electrons will be ejected*"
Inverse (v)—Indeterminate	A *verbal* discussion suggesting that wavelength is inversely related to KE or frequency.	"*As the wavelength decreases the* KE *increases*" *OR* "KE is inversely proportional to wavelength"
Linear (KE vs f)—Indeterminate	Any explicit mention (symbolic or verbal) that kinetic energy and *frequency* are directly proportional	$\mathrm{KE} \propto f$ OR "*the kinetic energy depends directly on the frequency*"
KE vs f (Plot)—Indeterminate	An explicit plot of KE versus frequency, whether it was their final answer or used as support for a different answer.	A sketch resembling plot (a)

generated in this discussion, suggesting it is sufficient to describe the major moves students might make.

The authors then met again to finalize the code book. In finalizing the code book, the authors analyzed the structure of the reasoning elements and coded these various tools as predominantly mathematical, predominantly physical, coordinated, or indeterminate in nature. The first three reasoning elements [*full equation, inverse (symbolic)*, and *asymptotes*] are considered more mathematical in nature; *cutoff (calculate)* is considered to be coordinated as it requires both the use of a mathematical calculation and a sufficient understanding of the physical system to know that there is a cutoff wavelength; *cutoff (verbal)* is considered physical in nature as student use of this reasoning element generally focused on the physical system, and nothing about the implicit mathematical formalism suggests there should be a cutoff. The remaining three reasoning elements [*inverse (v), linear (KE vs f)*, and *KE vs f (plot)*] are indeterminate in nature as they can, but do not necessarily, involve either mathematical, physical, or coordinated reasoning. The final code book, including descriptions and examples of the reasoning elements, is shown in Table I.

Coding the presence or absence of specific reasoning elements is necessarily reductionist. We do not claim that the presence of any one of these can capture the full nuances of student reasoning, or that a single reasoning element is always sufficient to code an associated MSM mode. Additionally, we do not believe that the presence or absence of any particular reasoning element is necessarily good or bad. However, again taking a resources perspective, we argue that the patterns of use of these reasoning elements (the frequency of use, and the clustered use of multiple reasoning elements) can provide a sense of the varied reasoning students are engaged in (as indicated by their MSM modes), as well as the dominant reasoning that leads students to draw a particular plot. We aim to categorize this reasoning as a mode of the framework, and show that reasoning elements can be used both to indicate an overall MSM mode and to highlight variations with these modes.

Following coding, two contingency tables were constructed relating the frequency of use of these reasoning elements to the plots drawn. The first table produces distributions of the reasoning elements present for students drawing each of the plot types. This association is largely descriptive in nature, providing a sense of the common reasoning involved in generating each plot based on the unique distribution of activated reasoning elements. It addresses the question: "if a student draws a particular plot, which reasoning elements are they likely using?" The second contingency table produces distributions of the plots drawn by students using a particular reasoning element. This view, while also descriptive, offers potential predictability, providing a sense of how likely a student is to draw a particular plot if they use a given reasoning element. Statistical hypothesis tests (Fisher exact and/or Pearson chi square) were conducted on these distributions where appropriate.

III. RESULTS AND ANALYSIS

The primary goal of this work is to show that the categorical framework, developed in the context of focus group interviews, can be productively applied to describe student reasoning and sense making using sparser data streams (written work and MC responses). In doing so, we also seek to provide an example of how this analysis is conducted, specifically the process of defining and using reasoning elements (both individually and in clusters) to validate the association of particular answers with MSM modes and to investigate variation within the MSM modes. Additionally, we provide preliminary evidence that reasoning and answer selection can be scaffolded by cueing particular reasoning elements, thus expanding the utility of the framework from a tool that is primarily descriptive to one that is also predictive of student reasoning. This first goal is accomplished through an in-depth analysis of student work on version B of the KE vs λ task from the semester 1 exam, while the second involves a comparison between student responses in semester 1 and semester 2 on versions B and C of the task, respectively.

To focus attention on these goals, our analysis of the semester 1 data revolves primarily around the four most common responses to version B of the KE vs λ task [plots (d)–(g)] which account for ~85% of student responses. The complete results of the semester 1 exam are shown in Table II. We briefly discuss plot (h), but ultimately lump these three responses and the single response for plot (b) into the other category. Likewise, our discussion of plot (a) is limited to the effect of overall framing (mathematical versus physical context) on engagement in the MSM modes. For those interested, a detailed discussion of the results of the pretest and homework for semester 1, including change-over-time through the exam, are provided in Appendix B.

TABLE II. Results of the semester 1 exam (version B). Because of the low counts, our analysis largely ignores plots (a),(b), and (h).

Plot	a	b	c	d	e	f	g	h	Other
Percent	4	1	0	12	15	39	18	3	8

A. Association of MSM modes with plots and written work

Before discussing the aggregate results from the semester 1 exam, characteristic reasoning accompanying plots (d)–(f) are qualitatively analyzed using the categorical MSM framework. These examples are representative of general student responses and specific student work for each of these four plots are shown in Fig. 10. By coding student responses for their MSM mode we establish a link between plot type (answer) and MSM modes. This link is explored more systematically using the finer-grained reasoning elements in later sections. We note that the analysis conducted here is consistent with the analysis of focus group interviews presented in prior work [8,22]. This consistency supports the use of the framework in analyzing written work. Additionally consistency between the reasoning present on the exam and in the focus group suggests that the framework captures aspects of sense making even in an exam environment with heightened stakes, where some sense-making activities may be subdued.

1. Plot (d): Msm-M

Plot (d) shows an inverse relationship, and while it is not explicitly negative there is no discussion of a specific cutoff wavelength (or frequency). As shown in Fig. 10, the reasoning accompanying plot (d) focuses on this abstract inverse relationship, citing without justification that KE $\propto 1/\lambda$ and verbalizing this mathematical relationship as "the longer the wavelength is the less KE it will produce." Though one might argue that this verbalization of the symbolic relationship counts as a physical interpretation, there is no reference to any particular physical entities (e.g., light as either a wave or photon, electrons, experimental setup), nor does this student explain a physical mechanism for "producing" KE. As is shown in Fig. 11 below, many students drawing plot (d) indicated the asymptotes of the plot in their work. However, this was generally only an indication that the axes *were* asymptotes, and student reasoning (as with that shown in Fig. 10) did not rely on this. Viewed together, this reasoning makes use almost exclusively of a symbolic, mathematical relationship as a tool for generating the plot. This is consistent with Msm-M reasoning. Though this plot does not completely describe the physical experiment it is a reasonable response based off a mathematical tool. While the tool itself is insufficient in its description of the phenomenon, use of this tool *does* represent productive, if incomplete, sense making.

2. Plot (e): Msm-M

Plot (e) is an accurate plot for the equation $\mathrm{KE}(\lambda) = hc/\lambda - \Phi$ without considering the cutoff wavelength. Importantly, as the question specifically asks for the kinetic energy, and the axes provided are labeled as KE, plot (e) allows for negative kinetic energy. As negative kinetic

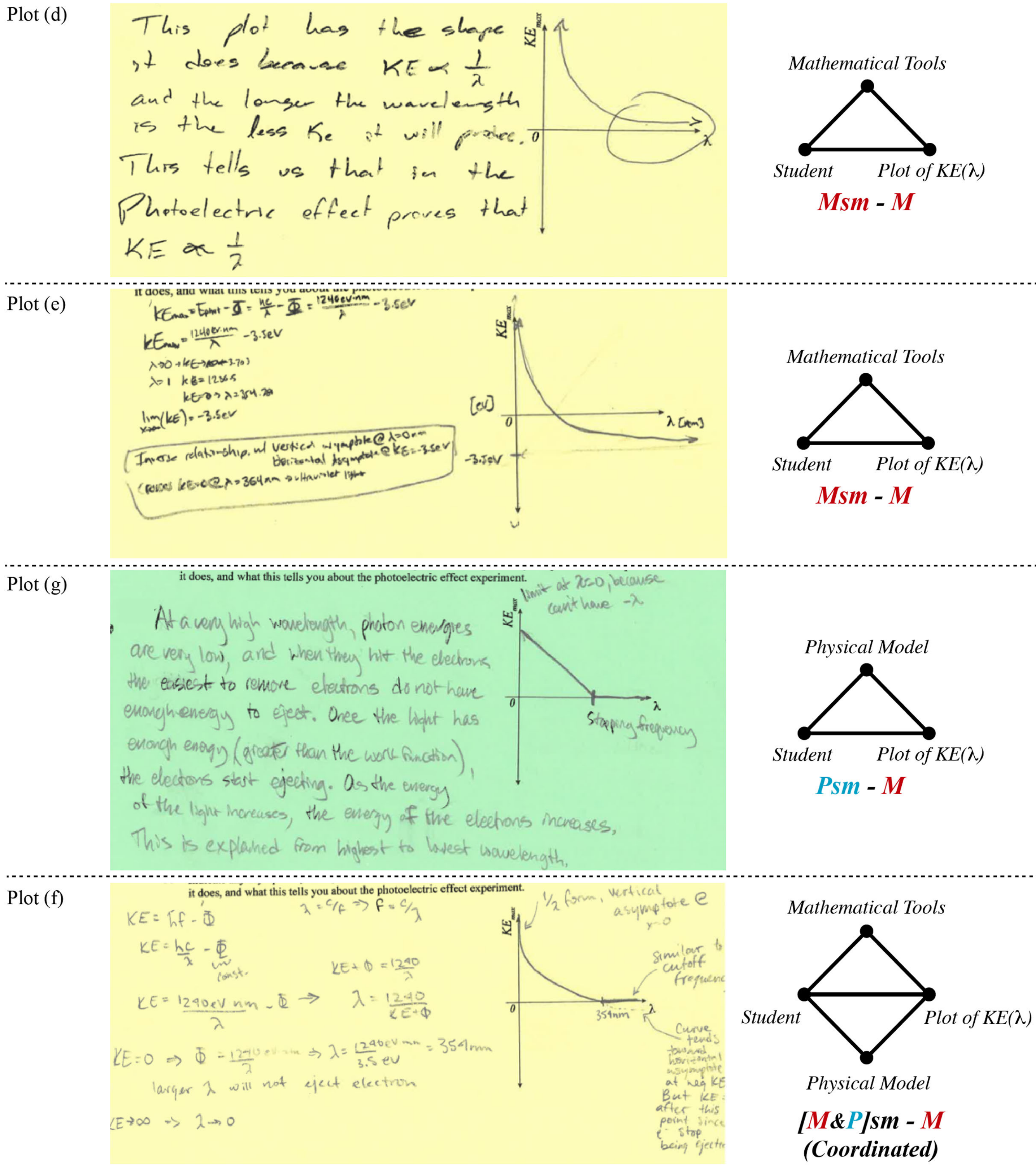

FIG. 10. Example student reasoning and associated MSM reasoning structure for plots (d)–(g). The color of the background (green or yellow) holds no meaning, these were simply the colors of the paper used for the two versions of the tests. The test versions were identical save for the ordering of some earlier questions. Note that the circle drawn on the tail of the graph in plot (d) was drawn by a grader, not the student.

energy is not a physically allowable concept (ignoring quantum tunneling) this is a strong indicator that students drawing this plot are *not* using a conceptual model of the physical experiment as a tool in their reasoning. As shown in Fig. 10 the reasoning supporting this plot draws predominantly on the mathematical equation written above.

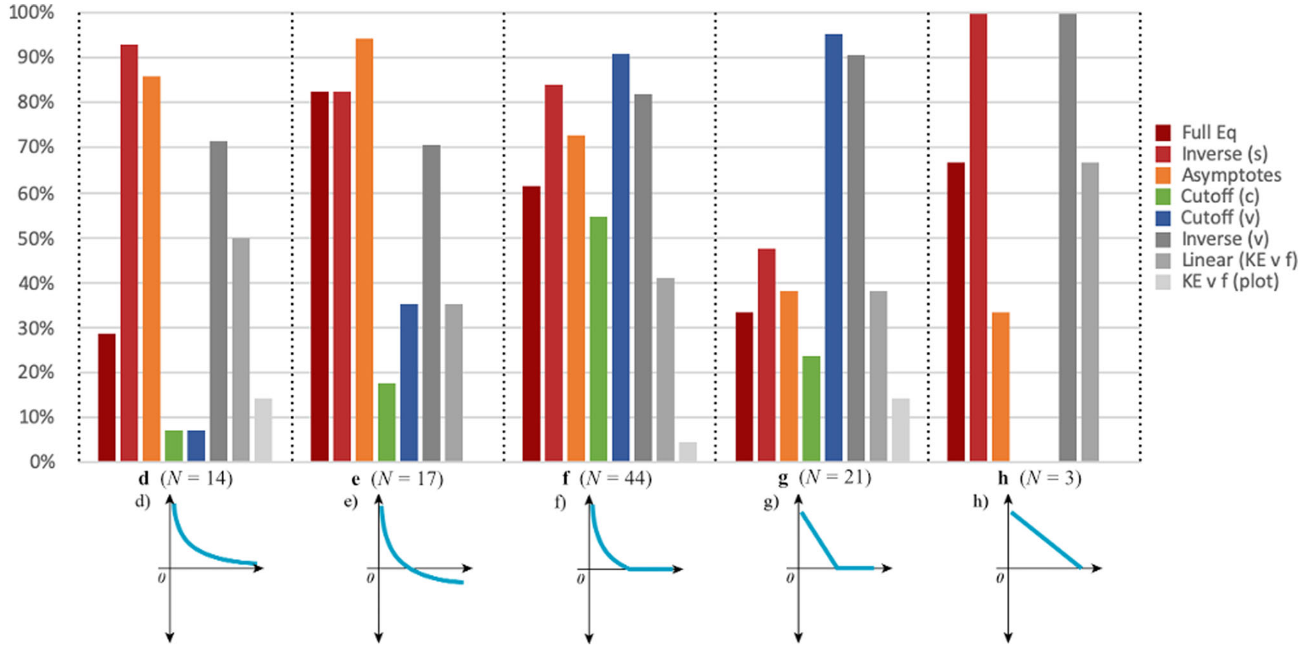

FIG. 11. The distribution of reasoning elements used by students who drew plots (d)–(h) on the semester 1 exam (version B of the KE vs λ task). The colors refer to our categorization of these tools as dominantly mathematical (red), physical (blue), coordinated (green), or indeterminate (gray). As *inverse (s)* is automatically included in *full equation*, the second bar is always greater than or equal to the first. The null hypothesis, that the distribution of reasoning elements is independent of plot, is rejected at the $p \ll 0.05$ level when comparing all five plots and the $p < 0.05$ level when comparing only plots (e)–(g).

In addition to an accurate plot of the equation, this student also examines (symbolically) limiting values for $\lambda = 0$, $\lambda = 1$, and $\lambda \to \infty$. Despite explicitly calculating the wavelength at which $\mathrm{KE} = 0$, the physical significance of negative kinetic energy does *not* override the mathematical formalism scaffolding the plot. Like plot (d) above, this plot is also consistent with Msm-M reasoning, though it shows an even stronger reliance on mathematical equations and processes as tools in sense making. As with plot (d), plot (e) does not accurately describe the physical phenomenon and even suggests a physically impossible negative kinetic energy, but this reasoning still shows productive sense making through accurate use of mathematical tools.

3. Plot (g): Psm-M

Plot (g) shows the wrong mathematical dependence below the cutoff wavelength but does explicitly show a cutoff wavelength above which the kinetic energy is zero rather than negative. As shown in Fig. 10, this plot was drawn with a strong consideration of the physical system in mind. This student mentions the relationship between wavelength and photon energies (verbally rather than with a symbolic inverse proportionality) and discusses the physical mechanism for electron ejection—the transfer of energy from the photon to the electron and the need to overcome the work function. They also correctly argue that as the energy of the light increases the energy of the electrons will too. Though there is no specific justification for this, they also state that λ cannot be negative, which is true only when λ is interpreted physically as the wavelength rather than a purely mathematical symbol. While this explanation relies on the inverse relationship between wavelength and photon energy, the relationship is discussed in a largely physical sense rather than in predominantly mathematical terms. This strong use of physical entities and quantities (photons, electrons, energy, work function) suggests this student is relying predominantly on a physical model of the phenomenon to generate this plot, which is consistent with Psm-M reasoning. Though the shape of the plot is incorrect below the cutoff wavelength it is true that the energy increases for shorter wavelengths. Thus, while not perfectly accurate mathematically, this plot accurately captures the main qualitative aspects of the photoelectric effect and shows productive sense making despite the incorrect mathematical mapping between an inverse relationship and the plot.[2]

[2]Anecdotally, this same physical reasoning was used by an expert in chemistry as she drew plot (g). In a course on Teaching and Learning Chemistry, J. D. G. was discussing this research project with the instructor following a class discussion regarding the photoelectric effect and quantization of energy. After J. D. G. posed version B of the KE vs λ task to her, the instructor began by considering the cutoff *frequency* below which there was no ejection, related this to a region above a cutoff wavelength and then stated that smaller wavelengths correspond to larger energies drawing a line rather than a curve. Thus, even for content experts, this Psm-M reasoning is compelling and productive when considering the physical phenomena without an explicit mathematical formalism.

4. Plot (f): Coordinated [M&P]sm-M

Plot (f), the correct answer, includes both the correct shape for the inverse relationship and an explicit cutoff above which the energy is zero. As shown in Fig. 10, the reasoning involved is generally more complicated than for plots (d),(e), and (g) and makes extensive use of *both* mathematical and physical tools. As with the reasoning accompanying plot (e), this student accurately generates the full equation $\mathrm{KE} = hc/\lambda - \Phi$, and considers limiting values of $\mathrm{KE} = 0$ and $\mathrm{KE} \to \infty$. However, unlike with plot (e), this student also relies on a physical model of the experiment to suggest that wavelengths above the calculated cutoff will not lead to ejection. While they indicate with a dashed line the negative values indicated by the expression (suggesting an asymptote at a negative KE), they use physical reasoning to override the explicit mathematical formalism, instead drawing a flat line that is "similar to [the] cutoff frequency." This reasoning is not fundamentally physical in nature, as with plot (g), or mathematical in nature, as with plot (e). Rather, this student makes use of *both* mathematical tools (the full equation plus asymptotes and limiting values) and physical tools (an understanding of the physical outcome of insufficient incoming energy). Though the plot is faithful to the equation, physical reasoning is used to override the (incomplete) mathematical formalism and create a plot that accurately describes the physical phenomenon. This complementary use of multiple tools is indicative of a *coordinated* reasoning structure, as discussed in the introduction and diagrammed in Fig. 2.

Though this particular explanation shows a coordinated use of mathematical and physical tools, it is possible to arrive at this plot through entirely mathematical reasoning (as in version A) or entirely physical reasoning that looks largely like the reasoning for plot (g) but with an accurate representation of an inverse.[3] Thus, while a student drawing (or selecting) plot (f) is not *necessarily* engaged in a coordinated reasoning structure, we argue both that it is likely that their reasoning is coordinated and that this coordinated reasoning structure is the preferred end goal of instruction showing fluid use of both mathematical and physical tools.

The analysis of written work for plot (f) is consistent with the analysis of the end of semester focus group presented in an associated paper [8]. In particular, while the focus group allowed for a finer grained analysis of student reasoning, the overall reasoning structure is the same. That is, the students in the focus group also developed a coordinated reasoning structure involving both the mathematical expression $\mathrm{KE} = hc/\lambda - \Phi$ and a physical model in which negative kinetic energy is not allowed and Φ determines whether or not the light has sufficient energy to eject an electron.

TABLE III. Percent engagement in the modes of the MSM framework for students drawing plots (d)–(g). These distributions are statistically significantly different at the $p \ll 0.001$ level using Fisher's exact test.

	Msm-M	Psm-M	Coordinated	N
Plot (d)	93%	0	7%	14
Plot (e)	100%	0	0	17
Plot (f)	9%	9%	82%	44
Plot (g)	5%	38%	57%	21

B. Semester 1 exam: MSM codes

While the characteristic reasoning of the last four subsections offers a preliminary link between plot type (answer) and MSM mode, it is expected that there will be variation in individual responses. To substantiate these associations, all student responses coded as plots (d)–(g) were coded for their MSM mode. Table III shows the percentage of students engaged in each reasoning mode for plots (d)–(g). These distributions are statistically significantly different at the $p \ll 0.001$ level using a Fisher exact test.[4]

In agreement with the analysis above, students that drew plots (d) and (e) are almost exclusively coded as engaging in an Msm-M mode. The one student who drew plot (d) using a coordinated MSM mode had reasoning more akin to that of students drawing plot (f), save that they said the kinetic energy would approach zero, rather than reach zero at a specific cutoff value. In contrast, plots (f) and (g) have a more diverse spread of MSM modes.

Responses for plot (f) included all three reasoning modes, with the Msm-M and Psm-M modes present equally but an order of magnitude less than the coordinated mode. This substantiates our analysis above that plot (f) can be associated predominantly with a coordinated MSM mode, while also capturing the (less likely) possibility that students are engaged primarily in either mathematical (Msm-M) or physical (Psm-M) reasoning. The greatest divergence between the analysis above and the results from the entire class occurs for plot (g). While plot (g) shows the greatest engagement in Psm-M of the four plots, this is not the majority mode. Rather, the majority of students drawing

[3]A relevant conceptual aspect of the inverse is that the rate of change is not constant. While we find this to be a particularly interesting conceptual interpretation that likely draws on an Msm-M mode, there was no indication that this reasoning was explicitly present for students. It could simply be that this was not externalized by students in this sparser data stream. We do not wish to imply that an accurate depiction of an inverse is trivial or does not involve Msm-M, merely that we do not have specific evidence of this reasoning here.

[4]The small counts in this contingency table, specifically the cells with values of zero, suggest the use of Fisher's exact test over Pearson chi square. As there are no marginal violations, the calculated p values are exact.

plot (g) were engaged in a coordinated MSM mode, and pure Msm-M was also present (though only for one student). This partially supports the analysis above, showing that plot (g) has a substantial association ($N = 20/21$) with Psm-M reasoning, but with the caveat that for the majority of students this Psm-M mode is coordinated with an Msm-M mode. Thus, while it is safe to say that students drawing plot (g) are engaged in a Psm-M mode, it is likely that they are engaged in Msm-M as well.

The MSM codes across the entire class are largely consistent with the specific examples of student reasoning presented above. This outcome suggests that the framework can be consistently applied to describe written work and establishes an association between MSM modes and specific plots (answers) in the context of the photoelectric effect. However, the high engagement in a coordinated mode for students drawing plot (g) suggests both that there is more varied reasoning associated with plot (g) and also that despite the large fraction of students engaged in a coordinated mode there is a difference in the finer-grained reasoning between students drawing plots (f) and (g). To explore the differences in reasoning (a sense of the mechanism of reasoning) for students drawing plots (f) and (g) and to further substantiate the association between MSM modes, written work, and MC answers, we now turn to a finer-grained analysis using reasoning elements.

C. Semester 1 exam: Associating reasoning elements and MSM modes

Figure 11 shows the distribution of reasoning elements that were present in written work for students who drew plots (d)–(h), e.g., 94% ($N = 16$) of the 17 responses coded as plot (e) discussed or calculated the asymptotes of their plot. Following the categorization discussed above, the mathematical reasoning elements [*full equation, inverse (s)*, and *asymptotes*] are plotted in various shades of red; the coordinated reasoning element [*cutoff (c)*] is plotted in green; the physical reasoning element [*cutoff (v)*] is plotted in blue; and the three indeterminate reasoning elements [*inverse (v), linear (KE vs f)*, and *KE vs f (plot)*] are plotted in shades of gray.

As each reasoning element *could* have been used by students that drew a particular plot, the maximum for each bar is 100%. This occurs only for plot (h), where all responses included a verbal discussion of the inverse dependence on wavelength; however, this is likely due to the small number of students that drew plot (h) ($N = 3$). Across the other four plots with larger N, there is no reasoning element present in *all* responses, though there are clearly dominant reasoning elements (present in over 80% of responses) for each. The clusters of these dominant reasoning elements suggest that there are consistent patterns of reasoning that lead to each plot, which we argue is aligned with the coded MSM modes above. The variations in these distributions suggest that these patterns of reasoning are different for each plot—with the exception of plot (h) which has too few responses to be statistically significant—corroborating our prior codes. The various distributions of reasoning elements by plot addresses the question "if a student draws a particular plot, what sort of reasoning are they likely engaged in?"

These distributions are statistically significantly different; that is, we reject the null hypothesis that the distribution of reasoning elements is independent of the plot drawn, at the $p \ll 0.05$ level when all five plots are included and when the low N plot (h) is removed. When considering just the most prevalent plots (e)–(g), which together account for almost 70% of total responses, we reject the null hypothesis at the $p < 0.05$ level. These p values were determined using Fisher exact tests with a simulated p value. For consistency Pearson chi squared tests were also conducted when considering only plots (d)–(g) and (e)–(g).[5] Each of these distributions will be analyzed individually below, looking specifically at the clusters of dominant reasoning elements.

Beginning with plot (d), the dominant reasoning elements are *inverse (s)* and *asymptotes*, which are both mathematical in nature (see Table I). *Inverse (s)* was present over three times more than *full equation*, suggesting that the general idea of an inverse relationship took precedence over a full mathematical formalism—a claim that is validated with the high use of *inverse (v)* and *linear (KE vs f)*. Together, this suggests that the majority of students that drew plot (d) followed similar reasoning to the example analyzed above in Sec. III A 1. Thus, this distribution of reasoning elements is consistent with Msm-M reasoning. We note that this distribution is qualitatively similar to that of plot (h), suggesting that students who draw plot (h) may follow similar reasoning to those that draw plot (d) except for a conflation of negative linear with inverse. While this may seem an obvious implication, another possible interpretation is that students drawing plot (h) held similar reasoning to those drawing plot (g) but failed to explicitly draw the flat portion after the cutoff. Our data suggest the former interpretation,[6] though we note that the small sample size is insufficient to validate this claim. As such, we will not consider plot (h) further in this analysis.

Plot (e) is similar in its distribution to plot (d). The three dominant reasoning elements are all mathematically focused, relying predominantly on *full equation*. Since writing the full equation automatically includes *inverse (s)*

[5]The small number of counts, particularly zeros, in some cells of these contingency tables lead to a chi warning, and suggest the use of Fisher's exact test. However, violations along both margins go against Fisher assumptions and lead to a simulated rather than exact p value.

[6]All students who drew plot (h) were coded as using an Msm-M mode, which further supports the claim that plot (h) more closely resembles plot (d) than plot (g).

it is not possible for *full equation* to appear more frequently than *inverse (s)*; however, the distribution for plot (e) is unique in that it is the only plot where these two reasoning elements are equally represented. This means that *all* instances of *inverse (s)* are associated with writing the full equation, providing strong evidence that plot (e) is associated with an Msm-M mode. This is consistent with the example of written work from Sec. III A 2, where the full equation is the primary tool used to mediate sense making. In addition to a reliance on the equation, these students also showed great attention to the *asymptotes* of their plot, producing a plot with a high fidelity to the symbolic formalism but ignoring the cutoff (an essential physical feature of a KE plot in a photoelectric context). It is also worth noting that the majority of the responses that include *cutoff (c)* and *cutoff (v)* are actually discussing the *crossing point*, as in the example from Sec. III A 2. While this crossing point is called out as a relevant feature of the graph, there is no indication that this point is interpreted as a cutoff wavelength, which further supports the claim that plot (e) is drawn using primarily Msm-M reasoning despite the increased use of these physical or coordinated reasoning elements.

The overall similarity between the distributions for plots (d) and (e), and the consistency between these distributions and the MSM modes above, offer strong support for the association of plots (d) and (e) with Msm-M reasoning. Despite the mutual association with a Msm-M mode, there are notable differences between the distributions of reasoning elements for these two plots. While the overall structure of the distribution can be associated with the MSM modes, giving a sense of the larger scale mechanism of reasoning (i.e., which tools are being used most frequently), the finer-grained differences highlight the variation in reasoning for students drawing these plots. This supports the utility of the categorical framework in labeling overall approaches by MSM modes and also indicates that not all instances of an MSM mode are identical or will lead to the same answer. This encourages the use of a reasoning element analysis when a finer grained understanding is desired.

The more mathematically focused reasoning elements that dominated responses for plots (d) and (e) play a substantially smaller role in the distribution for plot (g), being used by less than 50% of these students. Instead, student reasoning is dominated by a discussion (though not a calculation) of the cutoff wavelength [*cutoff (v)*] and a verbal (rather than symbolic) explanation of the inverse relationship between wavelength and frequency or kinetic energy [*inverse (v)*]. This cluster of dominant reasoning elements is in line with the example for plot (g) in Sec. III A 3 above, and so is suggestive of a dominant Psm-M mode. However, while this Psm-M cluster of reasoning elements dominates, there is still substantial use of the three mathematical reasoning elements, which is suggestive of more diverse reasoning than pure Psm-M. As was discussed above, there is strong evidence that students drawing plot (g) are predominantly engaged in a *coordinated* ([M&P]sm-M) reasoning mode, which is consistent with the substantial use of the mathematical reasoning elements.

From an inspection of the reasoning elements one might assume that the distribution for plot (e), in particular the non-negligible use of *cutoff (v)*, suggests that at least some of these students are engaged in a coordinated mode. The MSM codes do not support this, as 100% of student explanations accompanying plot (e) were consistent with an Msm-M mode. Rather, for students drawing plot (e) the *cutoff* reasoning elements are more akin to the *asymptotes* reasoning element, indicating attention to relevant points on the graph but with no indication that these students view the crossing point as having relevant physical meaning. While the cutoff codes are categorized as being either physical or coordinated in nature, their use by students drawing plot (e) is suggestive of a bidirectionality between reasoning elements and the MSM mode. In particular, that the use of clusters of reasoning elements can be associated with MSM modes but also that the MSM mode informs how these reasoning elements are used.

Where plots (d), (e), and (g) have no more than four reasoning elements present over 50% of the time [and for plot (e) all instances of *inverse (s)* come from *full equation]* the distribution of reasoning elements for plot (f) shows six of the eight reasoning elements present in over 50% of responses. This suggests that students who arrive at the correct answer draw on multiple reasoning elements as they generate and make sense of their plots. While one interpretation of this is that good students are simply "good" and give well-reasoned explanations, we argue for a more nuanced interpretation: while the majority of these reasoning elements are present in a majority of responses, there is still variation among them (not all elements are present equally), which suggests that there are multiple approaches to the correct answer. Additionally, only three reasoning elements are present in over 80% of responses for students selecting graph (f), which suggests that it is not simply good (comprehensive) students covering all their bases.

It is interesting to note that for plots (d), (e), and (g) there is a notable distinction between the use of reasoning elements *inverse (s)* and *inverse (v)*, however plot (f) has roughly equal usage of verbal or conceptual and symbolic argumentation around the inverse dependence on wavelength. The similarity in frequency between symbolic and verbal descriptions of the inverse, the high use of a majority of reasoning elements, and the substantial use of *cutoff (c)* all suggest that these students are largely engaged in a coordinated MSM mode, in line with the example analyzed in Sec. III A 4 and the MSM codes from Sec. III B. We note that the cluster of mathematical reasoning elements for plot (f) is more similar to that for plot (g) than plot (e)—that is, use of *inverse (s)* is not entirely due to use of *full equation*, and *asymptotes* is present comparatively less. This is

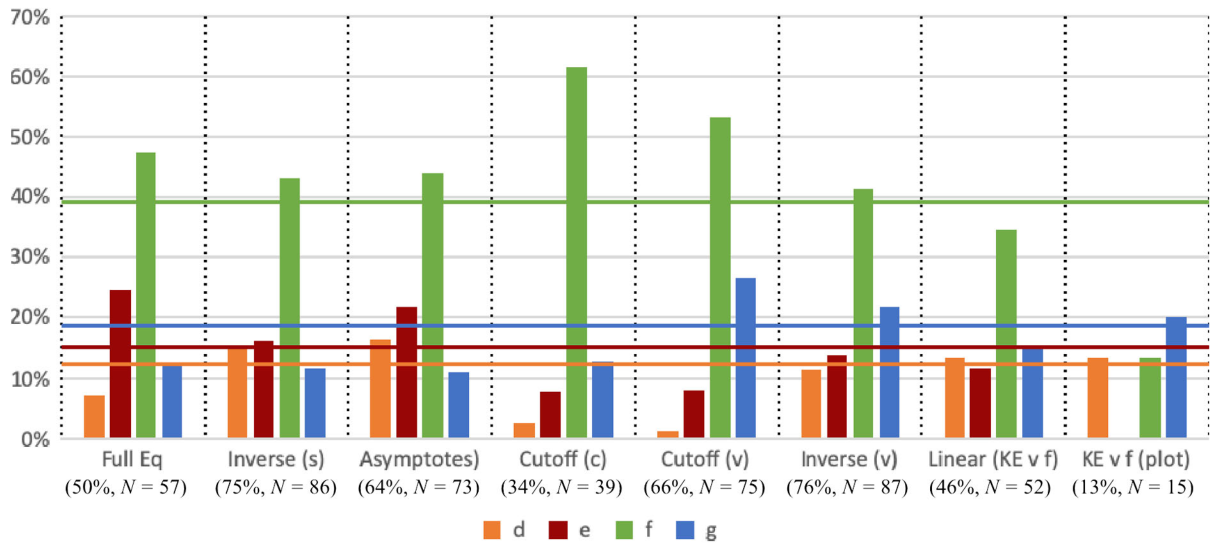

FIG. 12. The distribution of plots drawn based on the use of a given reasoning element. The horizontal lines show the percentage of overall responses corresponding to each graph, e.g., 39% of students drew plot (f)—the correct answer. These plots do not necessarily add up to 100%, as plots (a), (b), (h), and other are excluded from this figure. The overall percentage of student responses (regardless of plot drawn) that included each reasoning element is shown below each distribution, e.g., *full equation* was present in $57/114 = 50\%$ of students responses.

suggestive that the use of the mathematical reasoning elements for students drawing plot (f) is more akin to that of students drawing plot (g) than students drawing plot (e), offering further evidence that plot (f) is associated with a coordinated MSM mode, rather than equal but individual engagement in Psm-M or Msm-M. The amplification of the mathematical reasoning elements compared to plot (g) also hints at a more substantial coordination between the Msm-M and Psm-M modes for students drawing plot (f). This is in agreement with the MSM codes above, suggesting that while plot (g) is associated with a coordinated MSM mode, this coordination is less balanced and relies more substantially on the Psm-M mode.

D. Semester 1 exam: Associating individual reasoning elements with plots (answers)

The distributions above link *clusters* of reasoning elements with both MSM modes and particular plots (multiple-choice answers), substantiating the association between the MSM modes and particular plots. To further solidify this descriptive association, and as a first step at considering a potentially predictive association that could be leveraged to influence student approaches, we now look at how *individual* reasoning elements link to particular plots and MSM modes. That is, we ask the question: if a student uses a particular reasoning element, what plot are they most likely to draw and can this be associated with a given MSM mode? To address this question, we present the data not in terms of the distribution of reasoning elements used to generate a given plot, but in terms of the percentage of plots drawn based on the use of a particular reasoning element, as shown in Fig. 12.

The vertical bars in Fig. 12 indicate the percentage of student responses coded as containing each reasoning element that went on to draw each of the four plots [(d)–(g)], e.g., of the 57 students that used the *full equation* reasoning element, 27 (47%) drew plot (f). The horizontal lines show the overall response rate for each plot, e.g., $44/114 = 39\%$ of all students drew plot (f). The sum of the bars for each reasoning element do not necessarily add up to 100%, as plots (a), (b), (h), and other are excluded. These data help us to see patterns of use in student reasoning and strengthen our claim that the use of specific reasoning elements can be associated with both an MSM mode and a particular answer (plot drawn). In key instances, discussed below, a Pearson chi square analysis was conducted to substantiate the claim that the use of a given reasoning element is associated with graph selection.[7]

To isolate whether a given reasoning element can be associated with any particular plot(s), we compare the distribution of responses (plots drawn) for each reasoning element to the overall use of that reasoning element across all responses. If use of a reasoning element were equally associated with any of the plots, i.e., it did not discriminate among plot choices, we would expect that the distribution

[7]This is similar to asking whether a die is "fair", i.e., reproduces the expected distribution of outcomes. For a fair die the expected distribution is uniform, here our expected distribution is the overall response rate for each plot.

of responses for that reasoning element would be identical to the overall distribution. We observe that this is roughly the case for *inverse (v)*, where the distribution of plots in Fig. 12 corresponds with the horizontal lines (the expected usage). In contrast, when a particular plot is drawn substantially more (or less) frequently than the average response rate—i.e., when the vertical bar differs substantially from its corresponding horizontal line—we take it as a strong signal that the use of the given reasoning element plays a relevant role in discriminating between the plots.

The reasoning elements that show the most discrimination from the expected distribution are *cutoff (c)* and *cutoff (v)*. Students using these reasoning elements are very likely to draw either plot (f) or (g), with ∼25% of responses that mention (but do not calculate) the cutoff drawing plot (g). For both of these reasoning elements, we reject the null hypothesis that the distribution of plots is independent of reasoning element use at the $p < 0.05$ level, suggesting that these reasoning elements show a strong association with the final plot drawn and the associated MSM mode discussed above. In particular, *cutoff (c)* is the reasoning element most strongly associated with drawing the correct answer—a notable association as this reasoning element was used relatively infrequently (used by only 34% of students) but by more than half of the students that drew the correct answer. This association is unsurprising, as calculating the cutoff requires use of the full equation *and* the conceptual knowledge that kinetic energy cannot be negative and/or that there is a cutoff frequency (wavelength) that determines whether or not electrons are ejected. As this reasoning element relies on both mathematical and physical tools, use of this reasoning element is a strong indicator of a coordinated reasoning structure, consistent with the results of Table III and our analysis above.

Use of the full equation reasoning element was coded for 50% of students; it is likely to be associated with plots (e) and (f) and is less likely to be associated with plots (d) and (g). We reject the null hypothesis at the weaker $p < 0.1$ level. While this is generally not considered to be statistically significant, it is suggestive that use of the *full equation* reasoning element has a discriminating effect, albeit a less stable one than the *cutoff* elements, which is consistent with our argument. Based on these distributions, cueing the *full equation* is likely to lead students towards plots (e) and (f), cueing *cutoff (v)* is likely to lead students towards plots (f) and (g), and cueing *cutoff (c)* is likely to lead students specifically towards plot (f). However, the weaker p-value and the smaller deviations from expectation suggest that prompts cueing the *full equation* reasoning element will be less discriminating for the correct answer (f) than those cueing for the *cutoff* reasoning elements.

Use of *inverse (v)* was coded for 76% of students (the highest frequency of any reasoning element) and shows approximately equal association with all four plots. This result suggests that while it plays an integral role in student reasoning, this reasoning element does not discriminate well among the different plots. The remaining reasoning elements [*inverse (s)* and *asymptotes*] show weak discrimination, with a slight cue towards plots (d) and (e), which are associated with Msm-M, and away from plot (g), which is associated with Psm-M. Though there is only weak discrimination among plots from these two reasoning elements, their use is consistent with our categorization of *inverse (s)* and *asymptotes* (as well as *full equation*) as fundamentally mathematical reasoning elements. This outcome suggests that cueing the use of these reasoning elements is likely to encourage Msm-M reasoning (which could lead to coordination), but since Msm-M (or coordinated) reasoning is associated with all four of these plots this offers limited discrimination in the final answer.

The associations between individual reasoning elements, specific plots, and MSM modes further supports the link between plots (answers) and MSM modes. Additionally, the strong discriminatory effect of the *full equation*, and *cutoff* reasoning elements suggest that prompts scaffolding the use of these reasoning elements can play a predictable role in student reasoning—specifically in their final answer and engagement in particular MSM modes. Below, as a preliminary test of this predictive association, we consider the effect of a direct (visual) cue for the cutoff that does not explicitly cue the use of the full equation.

E. Semester 2 exam: effect of visual cueing for the cutoff

In version B (see Fig. 7), implemented on the semester 1 exam, we attempted to scaffold engagement in a Psm-M mode in drawing the graph through a calculation of a negative kinetic energy. However, for many students there was a mismatch between their responses to the calculation in part i and the plot drawn in part ii. In part i, only 54% of students clearly demonstrated reasoning that suggested they had a working physical model of electron ejection that was cohesive with their mathematical reasoning. More specifically, over 50% (9/17) of the students who drew plot (e) in part ii, which allows for negative KE, had explicitly stated in part i that the energy would be zero when their calculation returned a negative value. Exhibiting a similar mismatch, 10% of the students that drew plots (f) [4/44] and (g) [2/21] in part ii, which show zero KE above an explicit cutoff wavelength, accepted a negative kinetic energy when calculated in part i.

This mismatch between parts i and ii on the semester 1 exam suggests that our attempted scaffolding of Psm-M was ineffective; that is, the *calculation* of a negative kinetic energy did not necessarily cue students into a physical interpretation of this result. Even when students recognized this inconsistency in part i, this understanding did not necessarily carry over into part ii. In retrospect, it was not surprising that this intended cueing was somewhat ineffective, as it requires a calculation, a sufficient understanding of the physical system to accurately interpret the (unphysical) calculated result, and the continuation of this logic into future problem solving—a process we have called *chaining* in prior work [8].[8]

The results of the previous section establish a link between particular reasoning elements and the plot drawn (and so the associated MSM mode). Specifically, we note that the cutoff reasoning elements are the strongest discriminators for the correct answer. To test these associations, we designed version C of the KE vs λ task, predicting that a direct, visual cue would more effectively scaffold use of the cutoff reasoning elements. Since use of the cutoff reasoning elements is associated with a Psm-M mode, we expect to see a distribution of responses that is skewed towards the plots associated with Psm-M or coordinated reasoning.

As discussed in the methods section, version C (see Fig. 8) presents a physical scenario, and asks students to choose between the five most common results from semester 1 [plots (a) and (d)-(g)]. While it is still possible to use any combination of the reasoning elements observed in the free response version, and so engage in any of the prior MSM modes, this prompt provides more scaffolding by directing student attention to three differences between the plots. In choosing between these plots, students could consider (i) should the plot be straight or curved, (ii) increasing or decreasing, and (ii) is there a cutoff or can kinetic energy be negative? These first two points suggest activation of *inverse (v)*—perhaps strongly associated with *inverse (s)* for some students—and the third point suggests activation of *cutoff (v)*. Additionally, the overall framing suggests a Psm-M mode of reasoning rather than the Msm-M mode framed in version A. While *inverse (v)* was deemed as necessary but largely nondiscriminatory, the primary difference is the stronger visual cue for *cutoff (v)*. As discussed above, use of *cutoff (v)* is strongly associated with Psm-M (or coordinated) reasoning and leads towards plots (f) and (g) and away from plots (d) and (e).

Results of the version C exam question are shown in Fig. 13. A slightly larger fraction of students selected plot (a) in semester 2 than in semester 1 (10% versus 5%). As

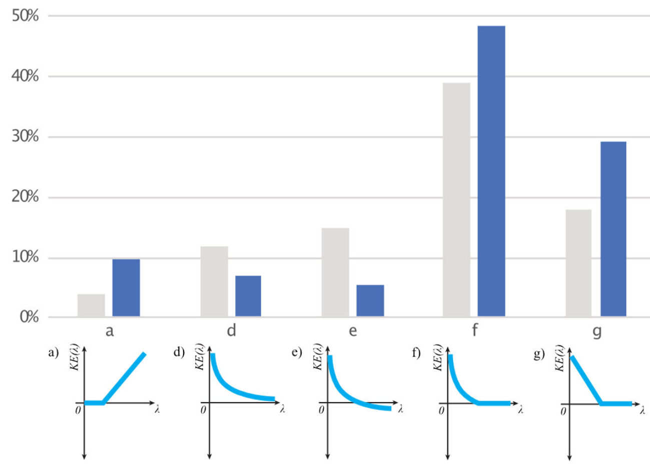

FIG. 13. Results for version C on the semester 2 exam (blue) and version B on the semester 1 exam (gray)—data for semester 1 are shown in Table II. The plot labels have been updated for consistency with the labels used throughout this paper.

there is no mathematical formalism given, it is unlikely that this is related to an interpretive difference between mathematical and physical functions, so—as in semester 1—these students were likely purposefully (if mistakenly) selecting a plot of $\mathrm{KE}(f)$ rather than $\mathrm{KE}(\lambda)$. In contrast to semester 1, very few students selected plots (d) or (e), suggesting a suppression of the pure Msm-M mode. Plots (f) and (g), associated with Psm-M and coordinated reasoning modes, make up the majority (~80%) of student responses here, and were selected substantially more than they were drawn in semester 1.

The shift away from plots (d) and (e) and towards (f) and (g) is in line with our predictions based on the use of *cutoff (v)* with no cue for *full equation*. This outcome provides evidence that cueing the cutoff wavelength through a visual representation may be a more effective cue for Psm-M and/or coordinated reasoning than asking students to perform a calculation. Admittedly, there are limitations to the comparisons that can be drawn between versions B and C, as they differ in format (free-response versus multiple choice) and required work (version B asks for both a calculation and a drawn plot while version C asks only for students to select a plot). However, our comparison is limited to the shift in plots chosen on version C from the plots drawn on version B. We argue that this shift is consistent with greater activation of the cutoff reasoning elements than was present on version B. Since neither version explicitly references the cutoff, this supports our hypothesis that a visual cue indicating the cutoff is more effective at scaffolding Psm-M reasoning than a calculation that gives a nonphysical result.

Additionally, representational form aside, these results substantiate the links between individual reasoning elements, MSM mode, and particular answers (plots). We

[8]It is worth noting that these data were collected from student work on exams. While exams were lower stakes than in many other courses, this is inarguably a time-sensitive and somewhat high stakes environment that may not have offered all students time to reflect on their work. We acknowledge this potential complication, however, our primary purpose in noting the inconsistencies between parts i and ii in semester 1 is to further support our association between the plots and MSM modes, not to make claims about a particular student's ability to sense make between problems. Both of these mismatches (rejecting negative kinetic energy when calculated but drawing regions of negative kinetic energy on their plot, or accepting negative kinetic energy when calculated but explicitly drawing plots with a cutoff) show that an understanding of the relevant aspects of the physical system was present for all of these students. The question then is not whether these students have the relevant resources, but whether they were activated during a given task.

argue that the distribution of responses to version C, shown in Fig. 13, is in line with large-scale engagement in a Psm-M and/or coordinated mode of reasoning, and thus take the shift in responses from version B as evidence that engagement in the various MSM modes can be scaffolded by appropriate activation of associated reasoning elements. The positive effect of a direct cue designed based on the MSM framework is suggestive that targeted curricula designed using this framework can scaffold both student reasoning (engagement in MSM modes) and the responses that they give.

IV. DISCUSSION AND CURRICULAR IMPLICATIONS

Where prior work [8,22] has applied the categorical MSM framework to analyze extended episodes of collaborative student reasoning, here we have expanded our analysis to multiple modalities of student work across a larger student population. Expansion to a larger N included sparser forms of student reasoning (individual written work and multiple choice answers) which are more commonly observed by physics educators. While it was not possible to see the nuanced changes in student reasoning evident in richer data streams (collaborative, think-aloud focus groups), it was still possible to code entire units of student work as being indicative of one or more modes of the MSM framework in a consistent fashion to the analysis of focus-group reasoning presented in our prior work [8]. This link between written work and multiple choice answers and MSM modes corroborates the utility of the framework in describing student reasoning across sample sizes and modalities of student work.

To further substantiate the link between MSM modes and particular plots (answers), a finer-grained analysis of student work was conducted looking at distributions of reasoning elements. The dominant clusters of reasoning elements were consistent with the coded MSM modes, validating the use of the MSM modes in describing the overall approach used to generate the various plots. These clusters of reasoning elements were also consistent with the analysis of exemplar written work, supporting the association of clusters of reasoning elements with a given plot, and so substantiating the link between MSM modes and particular answers. This provides strong evidence that multiple-choice answers can be indicative of specific MSM modes.

While the similarities between clusters of reasoning elements substantiates the association of multiple choice answers with MSM modes, variations in these clusters between plots associated with the same MSM mode provides more detailed information about differences in these MSM modes. For example, while plots (d) and (e) are both associated with Msm-M—and both draw predominantly on the mathematical reasoning elements *full equation, inverse (s)*, and *asymptotes*—the different frequencies of use of these three reasoning elements indicate a more sophisticated approach taken by students drawing plot (e) than drawing plot (d). Likewise, while plots (f) and (g) are both associated with a coordinated MSM mode, the distribution of reasoning elements is consistent with a more balanced coordination for plot (f) and a weaker coordination and larger reliance on the Psm-M mode for plot (g).

The agreement between the finer-grained coding of reasoning elements and larger-scale coding of MSM modes suggests that the categorical framework is a consistent and accurate tool for describing student reasoning across modalities of student work (focus group data, written, and multiple choice responses). Based on the analysis conducted here, we argue that the MSM framework can be a useful tool in validating (or invalidating) the association of multiple-choice answers with particular reasoning modes. This supports the use of the framework by both researchers and educators as a potentially useful tool for the analysis of student reasoning and performance on course artifacts.

Coding student work both for specific reasoning elements and the overall MSM mode also allows for discrimination between incorrect or partially correct answers. For example, in comparing between the incorrect answers (e) and (g) there is no *a priori* reason to believe that one is better or worse than the other. In fact, one could argue that plot (e) is indicative of "better" reasoning as it shows a more faithful understanding of the mathematical formalism that might also suggest a better understanding of the physical system. However, our analysis both of written work and the distribution of reasoning elements suggests that plot (e) is entirely based on Msm-M reasoning and does not show a complete understanding of the behavior of the physical system. On the other hand, plot (g) shows greater diversity in reasoning and suggests a greater understanding of the physical phenomena. In particular, a substantial fraction of student responses associated with plot (g) were indicative of a coordinated reasoning structure. This analysis suggests that plot (g) is more likely to be associated with the fluid use of multiple tools (an important aspect of mathematical sense making) than plot (e), and so could be seen as a better incorrect answer.

While not dwelt on here, the association between multiple choice answers and MSM modes suggests that the change in student responses over time can be used to measure student learning both in terms of canonically correct knowledge (content understanding) and also engagement in sense making. For example, between the pretest or homework and the exam in semester 1 there was a substantial shift away from plots (a)–(d) and towards plots (e)–(g), see Fig. 14 in Appendix B. This shift indicates broader engagement in more sophisticated approaches to this problem, whereas the shift towards plots (f) and (g) in particular indicates greater engagement in coordinated MSM modes and a more substantial use of multiple reasoning elements. While a measure of "correctness" is certainly a relevant measure of learning, this analysis is also suggestive of specific instructional interventions. For example, a student who has drawn plots (d) or (e) is likely engaged in purely Msm-M reasoning, and so a prompt to consider the physical system—specifically energetic requirements of electron ejection—could be a more productive approach than returning focus to a more detailed exposition of the full equation.

In addition to the descriptive utility of the framework, the association of individual reasoning elements with specific plots and associated MSM modes suggests a predictive ability. From semester 1 we see that the *full equation, cutoff (v)*, and *cutoff (c)* reasoning elements discriminate among the plots selected. Specifically that they likely lead students to plots (e) or (f), (f) or (g), and (f), respectively. The results of semester 2 suggest that the use of a given reasoning element is dependent on the representational form of the cue [42], and also that when effectively activated these reasoning elements show a predictable association with both particular plots (answers) and also with particular MSM modes.

Furthermore, the mismatch between responses to parts i and ii on the semester 1 exam indicate that engagement in a particular mode of MSM can inform the in-the-moment use of specific reasoning elements. For example, almost 50% of students who drew plot (e) on part ii, which is an accurate plot of the mathematical expression but allows for negative kinetic energy and so is not an accurate description of the physical system, had explicitly stated in part i that negative kinetic energy was not possible and so the electron would not be ejected. Likewise, several students accepted a negative kinetic energy in part i but went on to draw either plot (f) or (g) using their understanding of the physical system to state that there would be no ejection above the threshold wavelength. In both cases, this mismatch indicates that these students possessed a relevant understanding of the physical behavior of the system, but that this reasoning was differentially activated in parts i and ii. This is suggestive of a bidirectional interaction between reasoning elements and MSM modes: while the reasoning elements used (either individually or in clusters) are indicative of MSM mode, it is likely that engaging in a particular MSM mode influences which reasoning elements are used and to what end.

This predictability offers support for an approach to instructional interventions that scaffold engagement in MSM through the targeted activation of various reasoning elements. This approach could encourage the use and coordination of MSM modes, leading towards both greater MSM competence and a deeper (and more canonically correct) understanding of the relevant physical system. The utility of the framework to both describe student reasoning and support particular forms of student sense making and answer selection provides a basis for the use of the framework in the analysis and design of curricula.

V. CONCLUSIONS AND FUTURE WORK

We have expanded the application of the categorical MSM framework from in depth focus group data both to a larger sample size and to other modalities of reasoning, in particular written work and multiple-choice answers. An association between MSM modes and particular answers in the context of the photoelectric effect was established and then validated with a finer-grained analysis of reasoning elements. This suggests that the framework is a useful tool, for both researchers and instructors, for describing both the nuances of extended student reasoning and the general structure of shorter written responses. It can also be used to (in)validate the association between multiple choice answers and MSM modes.

This work was conducted in the context of the photoelectric effect, and in the main text and the Appendices we present a detailed discussion of the development of a novel task that links mathematical and physical reasoning and student responses and associated reasoning on this task. In Appendix A we include a discussion that explores potential conceptual issues regarding this task and show how it functions as a photoelectric shibboleth—a novel application of the reasoning associated with Corinne's shibboleth [7,37] that highlights differences between the interpretation and use of formal mathematics in the contexts of math and physics. In Appendix B we provide a detailed discussion of student reasoning on these tasks, and provide an example of how the framework might be used to track change-over-time in student reasoning.

In addition to the framework's utility in describing student reasoning, the association of individual reasoning elements with specific answers supports the potential predictive power of the framework. This predictive association was tested with a modification to the KE vs λ task in semester 2 that attempted a more direct form of cueing. The results of semester 2 provide preliminary validation of this approach, which could be employed to design curricula that scaffold the use of particular reasoning elements. These reasoning elements can be determined from an analysis of student work, as was done here, or could be predicted from assumed (or desired) reasoning.

In other work [44], we provide preliminary examples of curricula based on this approach. Future work will expand this pilot study to analyze existing curricula for the reasoning structures (MSM modes) it offers students the opportunity to engage in. We believe applying the framework in this way provides a tool for faculty or instructors (and curricular designers) to see the kinds of reasoning structures—i.e., MSM modes and use of specific (clusters of) reasoning elements—they are promoting. In addition to the analysis and design of curricula, future work will also test the efficacy of these approaches against more traditional (less scaffolded) curricula.

ACKNOWLEDGMENTS

The authors are grateful for discussions with and feedback from our collaborators in the PER group at the University of Colorado Boulder (in particular Steven J. Pollock, Bethany Wilcox, and Jessica R. Hoehn) as well as our collaborators at the University of Maryland. This material is based upon work supported by the National Science Foundation Grants No. DUE 1625824 and No. DMR 1548924. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

APPENDIX A: THE KE VS λ TASK

Here we detail the development of the three versions of the KE vs λ task, in particular the distractors in version A. In addition, we discuss several anticipated issues (both conceptual and procedural) with the mathematical formalism and expand our discussion of the photoelectric shibboleth.

As stated in the main text, version A (shown in Fig. 6) asks students to make an algebraic substitution to a given piecewise function and select the plot of the result. The five distractors were designed based on preliminary results from prior focus group studies and several assumptions of possible student difficulties based on this algebraic substitution. The intention was for plot (f) to be the only correct answer, with a shifted, inverse function before the cutoff and a constant value of zero after the cutoff. However, as discussed above, plot (a) is also a valid response for students that treat $f(x)$ as a mathematical function rather than a physical function. This will be discussed more below, but we briefly discuss the other distractors first.

Plots (b) and (c) retain the cutoff for small x and show either a *negative* linear graph or an inverse graph, respectively, for values above the cutoff. These distractors were chosen to appeal to students struggling with the idea of an inverse, as they might not consider the inversion of the bounds and could conflate inverse and negative proportionalities. This is in line with the work of Kwon *et al.* [36] who investigate some of the many difficulties students have with inverse functions and proportional reasoning. Plot (d) shows a pure inverse function, intended to capture students focused primarily on the inverse relationship between the variables, while plot (e) is an inverse function shifted down, effectively ignoring the piecewise nature of the function. Plot (e) is analogous to the plot of equation (1) and shown in Fig. 5; i.e., it is a correct plot of the primary functional dependence when the cutoff is implicit rather than explicit.

Plot (a) is a plot of the initial function [in this case $f(x)$] without making the change of variables. It was intended to be an incorrect option that appealed to students who simply chose the plot of the given function. Though we did not initially consider it in designing the task, plot (a) is a valid response due to the lack of physical context. It is analogous to the mathematician's answer to Corinne's shibboleth.

The standard version of Corinne's shibboleth gives the temperature of a rectangular metal slab as $T(x, y) = k(x^2 + y^2)$ and asks for the function $T(r, \theta)$. There are two "correct" answers to this question, that tend to distinguish between mathematicians and physicists. As Redish argues, a physicist sees T as a *physical function* (T_p) that represents a particular quantity (in this case the temperature) in whatever coordinates are specified. This suggests that (x, y) are Cartesian coordinates and (r, θ) are polar coordinates, in which case $T_p(r, \theta) = kr^2$. On the other hand, the mathematician sees T as a *mathematical function* (T_m) that represents a particular functional relationship between (an operation on) arbitrary quantities. Interpreted this way, x, y, r, and θ are dummy variables that hold no particular meaning and the mathematical function T specifies the operation "sum the squares of the two variables and multiply by k" giving $T_m(r, \theta) = k(r^2 + \theta^2)$.

Version A of the KE vs λ task involves a "pure-math" context, presenting the piecewise function $f(x)$ and asking for a plot of $f(y)$ given the relationship $x = D/y$, as shown in Fig. 6. Because of the lack of physical context and our choice to use generic variables x and y, which are often "meaningless" in a mathematical context, it is reasonable that many students interpreted $f(x)$ as a mathematical function. In doing so, the stated relationship between the variables is irrelevant and plots of $f(x)$ and $f(y)$ would look identical. Thus, plot (a) is the mathematician's answer to this photoelectric shibboleth, which (like Corinne's shibboleth) highlights how physical reasoning can override the traditional grammar of mathematical formalisms.

Beyond theoretical conjecture, this analysis was validated by the large number of students that selected plot (a) on both implementations of version A; see Appendix B, and a private (online) conversation a student had with J. D. G. In an email, this student said: "*If linear operator $f(x)$ is defined as $Ax - B$, then $f(y)$ will be defined as $Ay - B$ instead of $A(D/y) - B$. From the mathematical point of view, what you are actually asking is: Please select the graph of $g(y) = f(x) = Ax - B$ (where $x = D/y$). We shouldn't define f as one function and use it as another function*."

Ultimately, version A requires students to select a *graph* —an object that is fundamentally mathematical in nature, though inarguably contains physical significance given appropriate context. All five of the distractors, plots (a)–(e), were designed based on assumed difficulties with the mathematical formalism (conflation of inverse and negative linear, failure to switch or include the bounds, etc.). Because of the form of the distractors and the lack of physical context, it is expected that this question cues students towards an Msm-M mode of reasoning, where the generation (selection) and understanding of the plot is based primarily on the use of mathematical tools and procedures.

On the homework, this task was repeated as above but with the addition of a follow-up question that suggested this formalism is a relevant description of some aspect of the photoelectric effect experiment. Students were asked to consider relevant parameters in the experiment (current, voltage, kinetic energy, work function, incident light, etc.) and explain what this formalism might represent. This follow up question provides a physical system and was intended to encourage Psm-M reasoning such that students would draw on relevant knowledge of the photoelectric effect to contextualize the mathematical formalism.

In the context of the photoelectric effect experiment, this formalism can be written either in terms of the maximum kinetic energy or the stopping potential V_0 (the external battery voltage required to stop the most energetic ejected electrons from making it across the gap between the capacitor plates). The stopping potential is, effectively, an experimental measurement of the maximum kinetic energy: $\mathrm{KE}_{\mathrm{max}} = |eV_0|$. In this expression, e is the charge of an electron, which can be either explicitly or implicitly negative. Based on the orientation of the battery, the stopping potential is conventionally a negative number, which can also be either implicitly or explicitly written. If *both* signs are either explicit or implicit then there is no explicit minus sign relating the kinetic energy and stopping potential; however, if the conventional battery orientation is reversed, or *one* of the quantities but not the other is made explicitly negative, the relationship is written $\mathrm{KE}_{\mathrm{max}} = -eV_0$. A growing body of research has shown consistent difficulties with signed quantities in mathematics and physics, including a general trend to interpret all implicitly signed symbols as positive [45–47]. Thus, it is reasonable to believe that the signs in the relationship between kinetic energy and stopping potential pose an additional difficulty for students in understanding the mathematical formalism when considering the physical context of the photoelectric effect. Additional difficulties in performing mathematical operations due to the physical context are consistent with the work of Shaffer and McDermott, who found increased difficulty with simple vector tasks in the context of a collision [48].

Version B (shown in Fig. 7) was a free-response question implemented on the first midterm exam of semester 1. The KE vs λ task was presented explicitly in a physical context, and the standard mathematical formalism was not given. The bulk of our analysis in the main text focuses on version B, as we were interested in triangulating across written work and associated multiple choice answers. Because of the free-response context we expected that not all responses would fit into the six (mathematically motivated) plots present in version A, but save for the notable addition of plot (g) (see Fig. 9) this was largely not the case. The two parts of this question were intended to probe student facility with the mathematical formalisms (both symbolic and graphical) and to establish whether or not students had a sufficiently robust understanding of the physical system to "override" the mathematical formalism when it gave an unphysical result (negative kinetic energy). In practice, as discussed in the main text, for many students there was a mismatch in responses to parts i and ii where students either argued that kinetic energy cannot be negative when calculated in part i but drew plot (e) which is explicitly negative or accepted a negative kinetic energy when calculated in part i but drew plots (f) or (g) based on their understanding of the cutoff frequency or wavelength.

To test the hypothesis that a graphical representation presented in a physics context will provide stronger cueing for the cutoff frequency or wavelength, a multiple choice, physics context version was developed (shown in Fig. 8). Version C presents a physical scenario, a repeated photoelectric effect experiment with varied wavelength, and asks for the corresponding plot of KE vs λ. As this is a multiple-choice format the task is to select rather than generate a graph, and so is an arguably easier task than version B and comparisons between them are limited. That said, the possible answers on version C were the most common answers drawn by students on version B, and the primary task, to determine an accurate plot of KE vs λ, is the same. We argue that the most important distinction between the two versions are the ways in which attention to the cutoff is cued. While neither version explicitly mentions a cutoff wavelength, in version B a calculation is required that gives a negative answer for kinetic energy and so encourages an interpretation of no ejection while in version C the option of negative kinetic energy or a cutoff is visually presented. Responses to these two versions provide, at minimum, a preliminary indication of the effect of a more direct cue for the cutoff.

APPENDIX B: PRETEST AND HOMEWORK RESULTS

In the main text, we were primarily concerned with exploring the utility of the framework in describing student reasoning and connecting answer choices with the MSM mode. With that focus, a detailed discussion of student responses to the pretest and homework was not necessary. However, the framework can also be used to track change over time. As an example of this process, and for completeness in our presentation and analysis, we discuss the results for the pretest (version A) homework (version A) and exam (version B) for semester 1, which are shown in Fig. 14.

Responses to the pretest and homework were collected for all students in semester 1. Two students dropped the course after responding to the pretest but before the second homework and their responses have been removed from the dataset, leaving 108 matched pretest-homework-exam responses. On the homework, student responses to the physics follow-up were coded based on the physical relationship students stated the given mathematical formalism represented—e.g. that the function $f(y)$ represented the

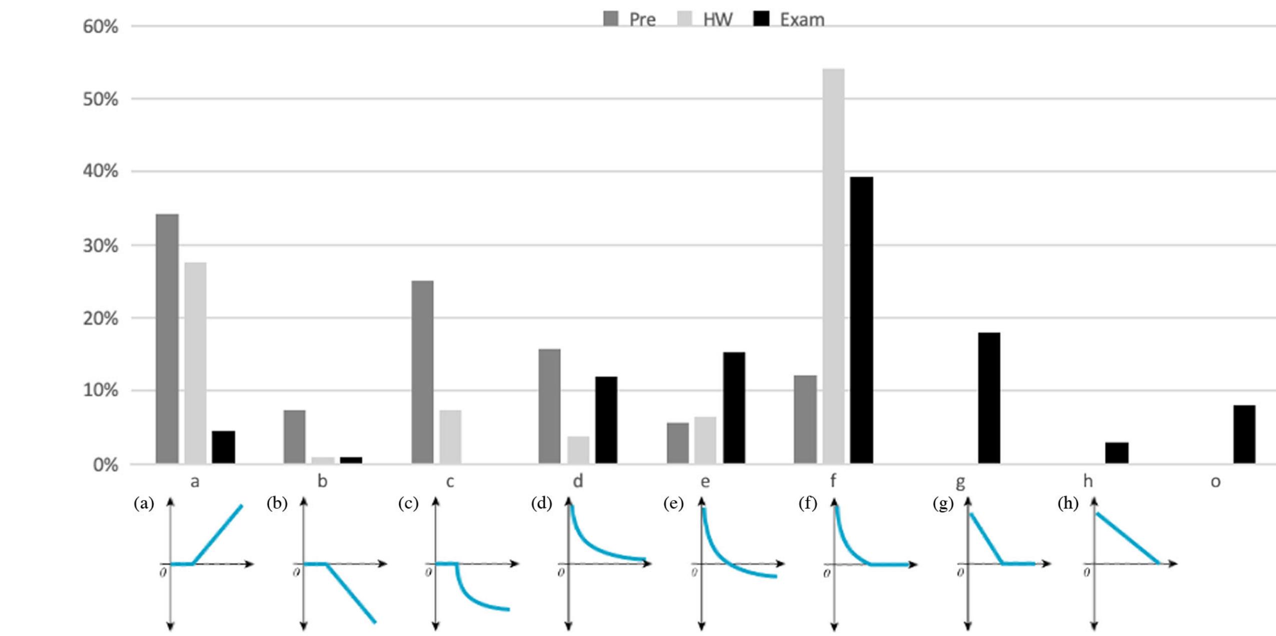

FIG. 14. Responses to the three implementations of the KE vs λ task for semester 1. Both the pretest ($N = 108$) and homework ($N = 109$) implementations were version A—though the homework included a free response, physics-context follow-up—and the exam ($N = 114$) was version B. Plots (a)–(f) correspond to the plots from version A shown in Fig. 6 and plots (g) and (h) (emergent from the free-response exam) are shown in the right of Fig. 9. Responses coded as o (other) make up less than 10% of the overall exam responses.

physical relationship KE(λ). There was a significant spread in these responses, but of the 100 students that gave responses to both parts of the homework all but 8 were captured by the 8 codes shown in Table IV.

While there is no majority response on the pretest, the most common response was plot (a), suggesting that 34% of students either selected the plot corresponding to the given function $f(x)$ without making the appropriate change of variables or that these students interpreted $f(x)$ as a *mathematical* rather than *physical* function—i.e., they gave the "mathematician's answer" to the photoelectric shibboleth. The second most common response was plot (c), suggesting students were aware of the inverse relationship for the *function*, but did not attend to the inversion of the cutoff; while the third most popular was plot (d) in which students attended to *only* the inverse nature of the function and completely ignore the cutoff. Only 12% of students chose the correct (for a physicist) answer, plot (f), on the pretest.

All six options were well represented in the responses, with no distractor chosen by less than 5% of students, which we take to show validation that the assumed difficulties led to sufficiently compelling distractors. Taken together, these results suggest that (i) this task poses difficulties on a purely mathematical level, regardless of its connection to the photoelectric effect and (ii) the majority of responses [59% across plots (c)–(f)] indicate an understanding that an inverse relationship appears as a curve rather than a line. It is also interesting to note that the least

TABLE IV. Results of the homework follow-up. Codes give the physical quantity and functional relationship students stated the formalism represented. Any zero counts are left blank. The codes KE, V_{stop}, and *current* indicate responses that either gave no dependent variable or cited a variable other than frequency or wavelength (e.g., position x). Responses coded as other referenced some other relevant quantity (e.g., the work function Φ or potential energy), those coded as "NA" chose a plot but provided no answer to the follow-up question.

	KE(λ)	KE(f)	KE	$V_{\text{stop}}(\lambda)$	$V_{\text{stop}}(f)$	V_{stop}	Current(λ)	Current	Other	NA	N
Plot (a)		18	1	1		1		3	3	3	30
Plot (b)			1								1
Plot (c)	1	1	1	2			2			1	8
Plot (d)	2		1				1				4
Plot (e)	2		1	1		1				2	7
Plot (f)	34	2	3	1	2	2	5	2	5	3	59
	KE or V_{stop}(something) $= 71\%$						current(something) $= 12\%$				

chosen distractor on the pretest, plot (e), becomes one of the three most common responses on the exam.

The homework implementation shows a large shift towards the correct answer; which is unsurprising due to the significant support that was available to students in conducting weekly homework. However, it is interesting to note that though plot (a) is still chosen by 28% ($N = 30$) of students, only 10 of these students had also selected plot (a) on the pretest. 18 of these 30 students also stated on the homework that the function $f(y)$ represents the kinetic energy as a function of *frequency*. This suggests that, on the homework, a majority of students selected plot (a) based on considering $f(x)$ as a mathematical function—indicating a lack of physical interpretation of the mathematical relationship between the variables x and y despite the explicit call to interpret these general variables in a physical context.

As plots (b) and (c) allow for negative values, something not *physically* possible for the kinetic energy, the shift away from these plots is suggestive of greater consideration of the physical system and engagement in physical reasoning (Psm-M). Of the 9 students who selected these plots there was a spread in their interpretation of what this function represents, including the current, the stopping potential, and the kinetic energy. While four students stated that these plots (which include negative portions) should represent the kinetic energy, this is perhaps a conflation with the stopping potential which *is* generally a negative number. On the other hand, six of these students stated that y represents the wavelength, suggesting a physical interpretation of the variables x and y, if not the function $f(x)$.

Of the 11 homework responses that selected either plot (d) or (e), which are increasingly accurate mathematical representations of the new function without considering the piecewise nature of the original expression, 7 indicated that $f(y)$ represented either the kinetic energy or the stopping potential, and 6 stated that y was the wavelength. Of the 59 students that selected the correct answer, 58% ($N = 34$) also correctly identified the function as $\mathrm{KE}(\lambda)$ and another 6 indicated that it was either the current or stopping potential as a function of wavelength. Across all responses, 71% of students indicated that the function represented either the kinetic energy or stopping potential, 12% indicated that it was the current, and 50% indicated that y represented the wavelength of light. On the whole, responses to the HW are indicative of learning in general and substantial physical interpretation of the functions $f(x)$ and $f(y)$ and the variables x and y (Psm-M reasoning).

Responses to the semester 1 exam question are again indicative of learning, as only 5% of students drew a graph resembling either plots (a) or (b), and no student drew a graph similar to plot (c). All 5 students who drew plot (a) on the exam had also selected plot (a) on the homework, and their responses are consistent with explicitly drawing a graph of kinetic energy versus *frequency*. While a plurality (though not a majority) of students drew a graph resembling the correct answer, there is a notable decrease compared to the homework (39% vs 54%). This is not particularly surprising as the format of the question is free response rather than multiple choice and the context is explicitly physical in nature rather than mathematical. Perhaps more surprising is that, despite having seen version A twice before with explicit solutions linking the function $\mathrm{KE}(\lambda)$ to the plots shown, 30% of exam responses include novel plots [(g), (h), and other]. Since other plots constitute only 8% of the overall responses, the eight lettered plots [(d)–(g) in particular] seem to effectively capture the dominant responses students give following instruction on the photoelectric effect.

The exam data have been analyzed in detail in the main text, so we will not repeat this analysis here. However, we do note that the change in student responses over time can be used to measure student learning both in terms of canonically correct knowledge (content understanding) and also engagement in sense making. For example, between the pretest, homework, and the exam there was a substantial shift away from plots (a)–(d) and towards plots (e)–(g). This shift indicates broader engagement in more sophisticated approaches to this problem, whereas the shift towards plots (f) and (g) in particular indicates greater engagement in coordinated MSM modes and a more substantial use of multiple reasoning elements.

[1] A. Gupta and A. Elby, Beyond epistemological deficits: Dynamic explanations of engineering students' difficulties with mathematical sense-making, Int. J. Sci. Educ. **33**, 2463 (2011).

[2] B. W. Dreyfus *et al.*, Mathematical sense-making in quantum mechanics: An initial peek, Phys. Rev. ST Phys. Educ. Res. **13**, 020141 (2017).

[3] K. T. Hahn, Student evaluative sensemaking on homework in intermediate mechanics, Master's thesis, Oregon State University, 2018.

[4] M. M. Hull, E. Kuo, A. Gupta, and A. Elby, Problem-solving rubrics revisited: Attending to the blending of informal conceptual and formal mathematical reasoning, Phys. Rev. ST Phys. Educ. Res. **9**, 010105 (2013).

[5] E. Kuo, M. M. Hull, A. Gupta, and A. Elby, How students blend formal and conceptual mathematical reasoning in solving physics problems, Sci. Educ. **97**, 32 (2013).
[6] E. Kuo, M. M. Hull, A. Elby, and A. Gupta, Assessing mathematical sensemaking in physics through calculation-concept crossover, Phys. Rev. Phys. Educ. Res. **16**, 020109 (2020).
[7] E. F. Redish and E. Kuo, Language of physics, language of math: Disciplinary culture and dynamic epistemology, Sci. Educ. **24**, 561 (2015).
[8] J. D. Gifford and N. D. Finkelstein, Categorical framework for mathematical sense making in physics, Phys. Rev. Phys. Educ. Res. **16**, 020121 (2020).
[9] L. Vygotsky, *Mind in Society* (Harvard University Press, Cambridge, MA, 1978).
[10] *Perspectives on Activity Theory*, edited by Y. Engestrom, R. Miettinen, and R. Punamaki (Cambridge University Press, Cambridge, England, 1999).
[11] W. M. Roth, Cultural-historical activity theory: Vygotsky's forgotten and suppressed legacy and its implication for mathematics education, Math Educ. Res. J. **24**, 87 (2012).
[12] T. O. B. Odden and R. S. Russ, Defining sensemaking: Bringing clarity to a fragmented theoretical construct, Sci. Educ. **103**, 187 (2019).
[13] B. R. Wilcox, M. D. Caballero, D. A. Rehn, and S. J. Pollock, Analytic framework for students' use of mathematics in upper-division physics, Phys. Rev. ST Phys. Educ. Res. **9**, 020119 (2013).
[14] O. Uhden *et al.*, Modelling mathematical reasoning in physics education, Sci. & Educ. **21**, 485 (2012).
[15] J. Tuminaro and E. F. Redish, Elements of a cognitive model of physics problem solving: Epistemic games, Phys. Rev. ST Phys. Educ. Res. **3**, 020101 (2007).
[16] B. Modir, J. D. Thompson, and E. C. Sayre, Students' epistemological framing in quantum mechanics problem solving, Phys. Rev. ST Phys. Educ. Res. **13**, 020108 (2017).
[17] D. N. Chari, H. D. Nguyen, D. A. Zollman, and E. C. Sayre, Student and instructor framing in upper-division physics, Am. J. Phys. **87**, 875 (2019).
[18] R. R. Bajracharya, P. J. Emigh, and C. A. Manogue, Students' strategies for solving a multirepresentational partial derivative problem in thermodynamics, Phys. Rev. Phys. Educ. Res. **15**, 020124 (2019).
[19] T. J. Bing and E. F. Redish, Analyzing problem solving using math in physics: Epistemological framing via warrants, Phys. Rev. ST Phys. Educ. Res. **5**, 020108 (2009).
[20] R. R. Bajracharya and J. R. Thompson, Analytical derivation: An epistemic game for solving mathematically based physics problems, Phys. Rev. ST Phys. Educ. Res. **12**, 010124 (2016).
[21] B. L. Sherin, How students understand physics equations, Cognit. Instr. **19**, 479 (2001).
[22] J. D. Gifford and N. D. Finkelstein, Categorizing mathematical sense making and an example of how physics understanding can support mathematical understanding, in *Proceedings of the 2019 Physics Education Research Conference, Provo, UT* (AIP, New York, 2019).
[23] A. A. diSessa, Toward an epistemology of physics, Cognit. Instr. **10**, 105 (1993).
[24] D. Hammer, A. Elby, R. Scherr, and E. F. Redish, *Transfer of Learning from a Modern Multidisciplinary Perspective* (IAP, Greenwich, CT, 2005) Chap. Resources, framing, and transfer.
[25] D. Hammer, Student resources for learning introductory physics, Am. J. Phys. **68**, S52 (2000).
[26] J. Minstrell, Facets of students' knowledge and relevant instruction, in *Research in Physics Learning: Theoretical Issues and Empirical Studies*, edited by R. Duit, F. Goldberg, and H. Niedderer (1991), pp. 110–128.
[27] A. Mortensen, Facet Innovations. Facet Innovations, https://facetinnovations.com.
[28] R. D. Knight, *Physics for Scientists and Engineers: A Strategic Approach with Modern Physics*, 3rd ed. (Pearson, Upper Saddle River, NJ, 2013).
[29] Photoelectric Effect. PhET. https://phet.colorado.edu/en/simulation/photoelectric.
[30] A. Sokolowski, Teaching the photoelectric effect inductively, Phys. Educ. **48**, 35 (2013).
[31] R. N. Steinberg, G. E. Oberem, and L. C. McDermott, Development of a computer-based tutorial on the photoelectric effect, Am. J. Phys. **64**, 1370 (1996).
[32] C. J. De Leone and G. E. Oberem, Toward undersanding student conceptions of the photoelectric effect, AIP Conf. Proc. **720**, 85 (2004).
[33] S. B. McKagan, W. Handley, K. K. Perkins, and C. E. Wieman, A research-based curriculum for teaching the photoelectric effect, Am. J. Phys. **77**, 87 (2009).
[34] E. Taslidere, A study investigating the effect of treatment developed by integrating the 5E and simulation on preservice science teachers' achievement in photoelectric effect, Eurasia J. Math. Sci. Technol. Educ. **11**, 777 (2015).
[35] M. A. Asikainen and P. E. Hirvonen, A study of pre- and inservice physics teachers' understanding of photoelectric phenomenon as part of the development of a research-based quantum physics course, Am. J. Phys. **77**, 658 (2009).
[36] Y. Kwon, A. E. Lawson, W. Chung, and Y. Kim, Effect on development of proportional reasoning skill of physical experience and cognitive abilities associated with prefrontal lobe activity, J. Res. Sci. Teach. **37**, 1171 (2000).
[37] T. Dray and C. A. Manogue, Functions. Oregon State University. https://bridge.math.oregonstate.edu/ideas/functions/.
[38] B. M. Wells, *Encyclopedia of Research Design* (SAGE Publications, Inc. Thousand Oaks, 2012), Chap. Naturalistic Observation, pp. 886–890.
[39] E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1997).
[40] F. Yates, Contingency tables involving small numbers and the χ^2 test, Suppl. J. R. Stat. Soc. **1**, 217 (1934).
[41] C. R. Mehta and N. R. Patel, A network algorithm for performing Fisher's exact test in $r \times c$ contingency tables, J. Am. Stat. Assoc. **78**, 427 (1983).
[42] N. S. Podolefsky, Analogical scaffolding: Making meaning in physics through representation and analogy, Ph.D. thesis, University of Colorado Boulder, 2008.
[43] S. Kapon and A. A. diSessa, Reasoning through instructional analogies, Cognit. Instr. **30**, 261 (2012).
[44] J. D. Gifford and N. D. Finkelstein, A framework for curriculum design to support mathematical sense making,

in *Proceedings of the 2020 Physics Education Research Conference*, virtual conference (AIP, New York, 2020).

[45] S. Brahmia and A. Boudreaux, Signed quantities: Mathematics based majors struggle to make meaning, in *Proceedings of the 20th Annual Conference of RUME* (2017), p. 1158.

[46] S. Brahmia and A. Boudreaux, Exploring student understanding of negative quantity in introductory physics context, in *Proceedings of the 19th Annual Conference of RUME, Pittsburgh, PA* (2016), p. 79, http://sigmaa.maa.org/rume/Site/Proceedings.html.

[47] J. Vlassis, Making sense of the minus sign or becoming flexible in "negativity", Learning Instr. **14**, 469 (2004).

[48] P. S. Shaffer and L. C. McDermott, A research-based approach to improving student understanding of the vector nature of kinematical concepts, Am. J. Phys. **73**, 921 (2005).