Student sensemaking about inconsistencies in a reform-based introductory physics lab

Jason M. May,[1,*] Lauren A. Barth-Cohen,[2,3] Jordan M. Gerton,[3]
Claudia De Grandi,[3] and Adrian L. Adams[2]

[1]*Department of Physics and Astronomy, West Virginia University, 135 Willey Street, Morgantown, West Virginia 26506, USA*
[2]*Department of Educational Psychology, University of Utah, 1721 Campus Center Drive, Salt Lake City, Utah 84112, USA*
[3]*Department of Physics and Astronomy, University of Utah, 115 South 1400 East, Salt Lake City, Utah 84112, USA*

(Received 25 April 2022; accepted 7 October 2022; published 17 November 2022)

There is growing interest in implementing reform-based lab courses in undergraduate physics that are student driven rather than instructor driven. In these courses, students develop and carry out experiments while simultaneously reasoning about their hypotheses, data collection procedures, collected evidence, and the relevant physics content. However, there is a limited understanding of how students reason in these types of labs. Using the theoretical framework of sensemaking, we examine qualitative observational data of undergraduates engaging in open-ended experimentation in reform-based introductory physics for life sciences labs. We examine the moment-by-moment details of students' sensemaking by focusing on a series of inconsistencies, specifically focusing on what the inconsistencies are *about* and what *moves* students enact during sensemaking to resolve them. We explore sensemaking about *conceptual* and *procedural* inconsistencies through a narrative case study analysis of a single group whose sensemaking is largely representative of the observed data corpus. We find that students engaged in sensemaking to resolve conceptual inconsistencies by juxtaposing hypotheses and evidence and critiquing and constructing scientific explanations, sometimes evoking elements of mechanistic reasoning. Comparably, we find that students engaged in sensemaking to resolve procedural inconsistencies by proposing and testing a series of causes towards modifying experimental procedures or apparatus. Overall, we find that student sensemaking about both types of inconsistencies is generally productive, given students' resolutions and experimental progress. Capturing detailed sensemaking about inconsistencies highlights the richness of students' reasoning processes in this understudied learning environment that is becoming more prevalent.

DOI: 10.1103/PhysRevPhysEducRes.18.020134

I. INTRODUCTION

There have been growing trends in physics education research (PER) to prioritize student agency and engagement in authentic experimental practices in new iterations of undergraduate introductory physics laboratory courses [1,2]. These reform-based lab courses are often built to provide students with greater intellectual and experimental freedom to direct their own experimental investigations [2,3], unlike traditional labs that emphasize more structured experimental procedures towards reaffirming previously learned content.[1] As a result, students often have increased opportunities to engage in physics reasoning processes during experimentation (Refs. [2,3,6,7]). In physics lab settings, student reasoning is often multifaceted, focusing concurrently on elements of experimental procedures, data

*jason.may@mail.wvu.edu

[1]We also point to contrasting cases of more historical reform efforts in undergraduate physics education, aimed at increasing students' conceptual understanding. These earlier reform efforts, commonly occurring in lecture courses, served as foundational motivation for highly scaffolded laboratory environments aimed at reinforcing lecture content, although these types of lab courses have generally been shown as ineffective at meeting this goal [4,5].

analysis, and physical concepts, all while students work to produce empirical results within a single lab experiment [6,8]. While student reasoning has long been a research focus of the PER community at large (Refs. [7,9–15]), there are fewer studies investigating the moment-by-moment student sensemaking processes *in situ* within introductory physics lab courses. This study presents a case study portrait of students engaging in reasoning processes to resolve various inconsistencies during experimentation in a reform-based introductory physics course. Using a sensemaking-based theoretical framework, we provide new insight into the moment-by-moment reasoning process of students in laboratory settings as they work to resolve both conceptual and procedural inconsistencies during experimentation. In this way, we capture the progression of students' sensemaking over time as they go through a complex process of identifying two types of inconsistencies and work to resolve them through various sensemaking moves. Thus, this study contributes new knowledge to the growing understanding of the nuances of student reasoning in this setting and, in turn, contributes to the pedagogical and empirical dialogue surrounding the development and implementation of these labs.

II. EXPERIMENTAL REASONING PROCESSES IN PHYSICS LABS

Parallel to the growing interest in implementing reform-based lab courses in undergraduate physics, recent research has focused on examining students' experimental reasoning processes within these labs. One line of research in PER has focused on students' singular reasoning processes related to individual experimental activities, such as troubleshooting [16], experimental design [3,17], representational interpretation [10], or measurement uncertainty [7]; sometimes, studies have an additional focus on demonstrating the effectiveness of new reform-based lab curricula. For example, Pollard and colleagues examined how recent course transformations in one physics lab course impacted how students reason about measurement uncertainty [7]. A different line of research has developed a series of large-scale conceptual frameworks of student lab experimentation [6,8,18,19], but these types of studies often omit moment-by-moment analysis of students' enacted moves while reasoning in lab settings. For instance, in analyzing the efficacy of the newly developed experimental physics cognitive task analysis (EPCTA; [8]), Holmes and Wieman showed that students in design-based labs have some opportunities to engage in rigorous experimental cognitive tasks, though their study was not designed to deeply explore the nuances of students' reasoning processes when they were engaged in these tasks [19]. Missing from this existing literature is a detailed observational analysis of students' experimental reasoning processes in labs that explicitly focuses on characterizing what students are reasoning *about* and what *moves* they enact during this reasoning. Especially pertinent is an empirical analysis of students' reasoning within labs where students are simultaneously reasoning about multiple aspects of the experiment—including their hypothesis, data collection procedures, collected evidence, relevant physics content—with high degrees of experimental agency, common in many newly developed introductory physics lab courses.

III. SENSEMAKING ABOUT INCONSISTENCIES IN REFORM-BASED LABS

To examine how students engage in scientific reasoning in this complex learning environment, we build on existing work on sensemaking in more scaffolded learning settings, such as structured tutorials [9] and clinical interviews [20]. Sensemaking is a "dynamic process of building or revising an explanation in order to 'figure something out'—to ascertain the mechanism underlying a phenomenon in order to resolve a gap or inconsistency in one's understanding" [21]. At a high level, we focus on two distinct elements of the sensemaking process: the process of "figuring something out" through explanation construction, and the nature and recognition of the inconsistency being resolved in the sensemaking process. This perspective on sensemaking is rooted in the overlapping resource theory and knowledge-in-pieces [22,23] and epistemic games theories [24,25]. With this perspective, we view sensemaking as an iterative process where students activate various "moves"—fine-grained actions residing within a larger reasoning process—while working to resolve gaps in their knowledge frameworks. Students use sensemaking moves, including building analogies, enacting mechanistic reasoning, and refining questions [20], to "figure something out" by first identifying the inconsistency in question then enacting various moves towards resolving the inconsistency and building a more coherent and structured knowledge framework. In this, the goal of this study is to identify the nature of students' experimental inconsistencies and describe the moves they enact while engaged in sensemaking to resolve them.

When considering what students may be sensemaking *about*, we expect that, in our studied physics lab setting, students often encounter a variety of inconsistencies, including *conceptual* inconsistencies when students' hypotheses do not sufficiently align with experimental evidence and *procedural* inconsistencies when students' experimental procedures create unexpected outcomes. While these are likely not the only forms of inconsistencies present, these two types of inconsistencies are likely prevalent in physics lab settings. As well, sensemaking about procedural and conceptual inconsistencies is synonymous with the procedural and conceptual knowledge continuum in mathematics education literature [26–29] and more broadly in PER literature [15,30–32]. More generally, inconsistencies often constitute forms of dissatisfaction for students that can necessitate a desire to engage in sensemaking to build satisfactory coherence [33].

Sensemaking about a conceptual inconsistency may involve students enacting several distinct moves. For instance, when working to resolve a conceptual inconsistency, students may reconsider their conceptual understanding embedded within the hypothesis and generate a more detailed explanation of physical concepts aligned with their experiment. For instance, Zwickl and colleagues documented how students recognize inconsistencies when comparing predictions of experimental phenomena (hypotheses) with measurement outcomes (evidence) during lab experimentation [6]. After recognizing this conceptual inconsistency, we expect students to iterate between collecting evidence, exploring and modifying their hypothesis, and assessing the alignment between hypothesis and evidence. Eventually, they might critique or renounce a prior hypothesis while finalizing the data analysis and building a new explanation of the experimental results. This explanation-building process can often involve "shopping for ideas" of related conceptual information [34] and an iterative process of construction and critique to enhance alignment to the experimental results [35]. This sensemaking process may also incorporate elements of mechanistic reasoning, as students are often tasked with exploring the physical mechanisms at play in an investigated phenomenon [36]. For instance, we expect that students' explanations may focus on identifying the entities, structure, and activities within a given system, doing so through analogies [37] or embodied action [38–40]. Resolution of the inconsistency in this process often involves a complete conceptual or mechanistic explanation of an investigated phenomenon that accounts for and utilizes experimental results to enhance the explanation. We expect that sensemaking about conceptual inconsistencies can be multifaceted and complex, offering students rich learning opportunities in physics lab courses.

Comparably, sensemaking about a procedural inconsistency in physics lab courses might center on the experimental procedures or the use of apparatus that lead to unexpected procedural outcomes. These inconsistencies are often identified by observing ongoing procedural tasks, whereby students see that a procedure or apparatus is not functioning as anticipated. A common example of this form of sensemaking is troubleshooting, a well-documented and important element of experimentation [1,16,41,42]. When making sense of procedural inconsistencies, students likely rely on distinct forms of knowledge, including system knowledge—an understanding of the structure and function of apparatus or software—and procedural knowledge—understanding of appropriate methods to utilize and adjust apparatus and software during experimentation [42]. Students use this knowledge to iteratively propose causes for the inconsistency, test whether these potential causes are the real issues, and then repair or adjust their apparatus or revise procedures to resolve the inconsistency [16,41]. Similar to making sense of conceptual inconsistencies, sensemaking about procedural inconsistencies often involves nonlinear and iterative student action, presenting challenges during experimentation but affording students valuable opportunities to enhance experimental skills.

There are some important distinctions and similarities in sensemaking about these two inconsistencies. There may be differences in the role of evidence and observation in the sensemaking process, the relevant knowledge students utilize to achieve resolution, and the timescale within the larger experimental process. Additionally, there might be instances of sensemaking in these lab courses that involve attributes of both forms of sensemaking described here. We focus on the students' perspective of their sensemaking but also recognize that it can appear different from the educator or researcher perspective. For example, if students' sensemaking involves adjusting apparatus or modifying their experimental procedures to resolve an inconsistency, the inconsistency, from the students' perspective, was likely viewed as procedural even if an outside observer may recognize an underlying conceptual inconsistency. Thus, we recognize that these two types of sensemaking can be intertwined, and we expect these two types of inconsistencies exist on a continuum rather than being binary opposites. Given this perspective on sensemaking, our goal is to examine the moment-to-moment detail of student sensemaking about inconsistencies in reform-based lab courses, focusing on characterizing the nature of the inconsistency in question and the moves students enact during their sensemaking.

IV. RESEARCH QUESTION

To examine undergraduate students' sensemaking in reform-based introductory physics lab courses, we focus specifically on identifying and characterizing the inconsistencies that initiate students' sensemaking processes and determining what moves students enact to carry out their sensemaking. To this end, we ask the following research question: What forms of inconsistencies are students in introductory physics lab courses sensemaking *about*, and what *moves* do students enact during this sensemaking to achieve resolution?

Here we present two instrumental focal cases of students' engaging in sensemaking about inconsistencies in a reform-based introductory physics lab course. Across both cases, we characterize the form of inconsistency under scrutiny and examine in detail the moves students utilize as they engage in sensemaking to resolve their inconsistencies. The research is explored within the context of an undergraduate introductory physics for life sciences (IPLS) lab course that is structured to provide student groups with experimental agency to develop and carry out experiments that explore the physical properties of biological phenomena. These results will contribute to a deeper understanding of the inherent complexity of students' experimental reasoning processes in introductory physics lab courses, working to fill a gap between two existing bodies of literature that, respectively, focus on analyzing student reasoning within single experimental processes and describing the nature of student experimentation through broad experimental frameworks.

V. METHODS

A. Study context and participants

Data come from a first-semester IPLS lab course at a large, research-intensive university in the western United States [43]. The course, along with its second-semester counterpart, was recently reformed in line with recent stakeholder recommendations (i.e., Refs. [44,45]) and the NEXUS/Physics curriculum [46] to prioritize students' open-ended scientific experimentation focused on investigating the physical properties of biological systems [43].

The student population, comprising 219 students, largely consisted of upper-division undergraduate students (86%) from life science disciplines (e.g., biology, kinesiology); most enrolled students (70%) stated on course surveys that they plan to attend medical school upon graduation. The student population had the following demographics: 68% White, 30% domestic students of color, and 2% international students; 47% female and 53% male [47].[2] For most enrolled students, this is the first physics lab course they have taken. The lab course is separate from its co-requisite lecture counterpart, with minimal overlap in content.

In the course, students work in groups of four (or three) in each multiweek lab investigation to (i) develop a research question, experimental design plan, and initial experimental hypothesis broadly aligned with the lab's guiding prompts, experimental apparatus, and software, (ii) conduct an investigation to study their chosen scientific phenomenon, (iii) develop a scientific argument consisting of a claim (result), experimental evidence, and detailed scientific explanation (reasoning) [48,49], and (iv) present their scientific argument to their peers, teaching assistants (TAs), and learning assistants (LAs) via whiteboard presentations and written lab reports. Lab investigations comprise either two or three lab sessions, each of which is 3 h long. Overall, student groups have roughly 3–5 h of in-class lab time during each multiweek lab investigation for formal experimentation, with the remaining time allocated for warm-up activities, whiteboard presentations, and writing lab reports. Student groups have a broad goal of developing and carrying out an experiment that they can use to study a scientific phenomenon related to guiding prompts in the lab documentation. Students consistently utilize their lab computers, computer software, and associated experimental apparatus (e.g., microscopes, pipettes, biological samples). More generally, the course is designed to limit direct instruction (either verbally or written) to provide groups with opportunities to develop and carry out experiments with minimal scaffolds. For example, in one lab investigation, students are prompted to "create an investigation that studies Brownian motion that can provide evidence and/or insight into how diffusion occurs"; students are also provided sample experimental trajectories, such as identifying ideal cellular environments (e.g., viscosity, temperature) for Brownian motion within a cell.

In this course, TAs and LAs have a unique role as they function as more knowledgeable others and are tasked to support students' own experimentation rather than direct students toward a particular outcome [50]. The TAs and LAs provide general guidance and feedback on groups' experimental progress and decisions but are trained to not provide direct instruction to student groups related to making experimental decisions or interpreting results. For this reason, TAs and LAs are often more involved with procedural matters, such as when experimental apparatus is exceptionally faulty, as compared to conceptual matters during experimentation. Furthermore, in this setting, students are provided with a guiding prompt to scaffold their experimentation, from which they work to develop a research question and experimental design; however, neither students nor instructors know the experiment's outcome beforehand, which is fundamentally different from traditional labs.

1. Researchers' roles in instructional context

Beyond our involvement in the present study as researchers, several authors were directly involved in the reform efforts and teaching of the studied IPLS course environment. The first and third authors led the course reform effort beginning in Fall 2017, focusing on the first-semester course with assistance from additional faculty and students, with the fourth author joining in Summer 2018 and leading the reform effort for the second-semester counterpart course. During the semester when this study took place, the first author served as a TA for several observed sections and groups, and the fourth author served as instructor of record for the course.

[2]We note that student demographic data come from the study institution's Office of Budget and Institutional Analysis [47]. The OBIA demographic dataset did not provide data on students' gender identity beyond a binary model of male or female, nor did it provide additional categorization of students' racial or ethnic background. We recognize that there were likely additional gender identities and racial and ethnic backgrounds present in the student population that were not sufficiently captured in this institutional data.

B. Data collection

Data were collected from in-person class observations. The first author collected screen capture data of student groups' computers and external video and audio data of students working at their lab workspaces. This data corpus was chosen to allow triangulation of multiple perspectives on student dialogue, gesturing, use of computer software, interactions with apparatus (e.g., microscope), and peer-peer and peer-facilitator interactions. The groups were selected by identifying all groups whose individual members all consented to participate in the research study. Thus, we did not collect data from groups where some but not all members consented. Overall, we collected observational data from thirteen student groups ($N = 38$) across four multiweek lab investigations over one semester.[3]

C. Data analysis

After data collection, the first author enacted an initial analysis by working from field notes to begin a first coding cycle [51] that focused on different actions students enacted while working with their data [52]. Working from the data and relevant literature [6,16], we made detailed notes, transcribed selections, developed a series of concept codes (e.g., "equipment or software troubleshooting"; [53]) that captured when students were working with experimental data, then iteratively applied these codes to selections from the data corpus while developing, revising, and refining the code book until saturation was reached. The research team met multiple times during the initial coding process to discuss emergent codes and enhance code book reliability and consistency. We then transitioned into the second cycle of coding, where we grouped the codes into two categories, those codes associated with direct experimental actions (e.g., "data collection") and those associated with supporting actions that are dependent on the direct actions (e.g., "procedural interpretation"). The first and fifth authors conducted an interrater reliability measurement with the refined code book and categorizations, reaching a Cohen's kappa value of $\kappa = 0.83$, indicating near perfect agreement [54], after a single round of code book definition revisions.

In the second analysis phase, the first and second authors worked from the entire coded data corpus to identify potential episodes of rich sensemaking. We started by examining the entire data corpus for episodes with a high density of codes, particularly clusters that contained many *conceptual interpretation* or *procedural interpretation* codes, as aligned with our goals and theoretical framework. We also examined the data for (i) when the students encountered an inconsistency during experimentation, evident by confusion or uncertainty about any number of aspects of the experiment, and (ii) when students were constructing or critiquing explanations or procedures. We also noted when the students were notably talkative, including within their groups, with other groups, or with instructors, as this can indicate more verbal evidence of their sensemaking.

Next, we reviewed these episodes as potential cases for further analysis. The first author wrote analytical memos [51] to capture the essence of student sensemaking, which included identifying potential inconsistencies, steps the groups took to resolve the inconsistency, and whether a resolution was ultimately achieved. In line with our theoretical framework, we also incorporated details about how students' sensemaking may have incorporated elements synonymous with troubleshooting or mechanistic reasoning. Once the memos were complete, we used them to categorize the cases as either procedural or conceptual inconsistencies more rigorously. That is, episodes where the inconsistencies mainly involved the experimental apparatus or student-driven procedure were categorized as procedural, and episodes in which the inconsistencies centered more on the mechanisms involved in the physical phenomena and students' understandings of it were categorized as conceptual. Furthermore, we identified common trends among student groups related to the sensemaking steps taken to resolve recognized inconsistencies. Then, the first and second authors reviewed the memos and episodes several times to identify a single student group whose experimentation involved rich sensemaking examples about procedural and conceptual inconsistencies. The entire research team met together frequently to discuss possible groups that best represented the data corpus during this selection process. In this way, we selected a single instrumental focal group using intensity sampling [55], from which we analyzed two different cases of sensemaking. This focal group's experimentation and sensemaking were representative of the larger data corpus, but this focal group is also unique in that it contained many inconsistencies and the students were rather talkative, both to each other and to the instructors, providing rich data to examine their sensemaking.

The first author then created a detailed transcript of the group's experimentation to facilitate further analysis, focusing on potential sensemaking episodes. Transcript conventions and techniques were chosen to help elicit student reasoning (see Table I; [56–58]). For example, we transcribed for intonation shifts, laughter, and pauses to direct us to notice possible student uncertainty or confusion, which we interpret as a potential mark of inconsistency. Similarly, we transcribed for overlapping speech and dialogue latching to help us notice nuances in the discussion.

To highlight the necessity of these transcript conventions, we provide a brief example from the data corpus. Consider the dialogue below, taken from early in the focal

[3]In this course, student groups were shuffled after each multiweek investigation. These rotations caused that some student participants were observed in multiple lab groups during the semester, while other students were only observed once.

TABLE I. Transcription conventions.

Symbol	Meaning
[…]	Overlapping speech
=	Latching; when one speaker's utterance is immediately followed or cut off by another speaker's utterance with no gap.
(0.0)	Indicates a pause in speech for a given length in seconds (2.0).
word	Underlining indicates vocal stress via pitch or amplitude. Shorter means lighter stress than longer.
wo:rd	Indicates an up-to-down intonation contour.
wo:rd	Indicates a down-to-up intonation contour.
wo:rd	Indicates an up-down-up intonation contour.
wo:rd	Indicates a down-up-down intonation contour.
wo:rd	Indicates a whole word up-shift in intonation from prior word. No midword shift.
wo:rd	Indicates a whole word down-shift in intonation from prior word. No midword shift.
.,?!	Indicates the usual punctuation intonation.
∘word∘	Indicates especially soft dialogue (e.g., whispering).
—	A dash indicates a cutoff in speech.
>word<	Indicates speech that is sped up or hurried compared to other speech.
<word>	Indicated speech that is slowed down compared to other speech.
( )	Indicates inaudible speech, according to the transcriber. Space included represents length of time compared to other dialogue.
⟨*text*⟩	Indicates transcriber's description.
@	Indicates laughter (each @ is a single pulse). When attached to word, indicates utterance with mild laughter.
h	Indicates breathiness or extension of syllable within a word (e.g., Oh*hhh*). Number of letters indicates length.

group's data collection, shown first without transcript conventions:

938	Lisa	Eh. You're close. Oh, there we go. Oh, look at that. Oh, that's really nice.
939	Mika	That, they are moving faster, right?
940	Lisa	Beautiful.
941	Mika	Or no?
942	Lisa	I don't know if they're moving that much faster.

This can be compared to the same dialogue with full transcript conventions:

938	Lisa	⟨Adjusting microscope dials to bring 500× sample into focus⟩ Eh*hhhhhhhh*. You're clo:se! Oh, there we go. Oh, look at that. Oh, that's really nice.
939	Mika	That—they are moving faster, [right?
940	Lisa	[Beautiful!
941	Mika	Or no?
942	Lisa	∘I don't know if they're moving that much faster.∘

In the latter case, the nonverbal description of student activity in line 938 provides experimental context for Lisa's statement that a reader may not identify in the original transcript. Likewise, the additional tonal stress (underlined transcript) in Lisa's dialogue may suggest excitement or confidence in her actions and observations. Finally, when comparing line 942 in both cases, we see that the inclusion of notation demarcating soft dialogue may prompt a reader to interpret Lisa's statement as containing uncertainty in her observations of microsphere speed, in addition to potential hedging ("I don't know if …") of Mika's conflicting observation in line 939 [59].

Using the new transcript in combination with the previously created memos, we identified seven episodes of the group sensemaking about procedural inconsistencies, each lasting under 30 min, and one episode of sensemaking about a conceptual inconsistency that spanned nearly 3 h across two lab weeks. Using the transcript, we annotated episodes of sensemaking about procedural inconsistencies, identifying the inconsistency involved, the sensemaking process taken towards resolution, whether resolution was achieved, and what explanation was offered for the resolution. Comparing the seven procedural episodes, we found that the group often iterated between generating potential causes for the procedural inconsistency and proposing, carrying out, and assessing procedures to test those same cases. In many ways, the nuances of this iteration provide new moment-by-moment insights into previous results studying experimental troubleshooting in physics lab settings [16]. The results present one of these episodes with rich dialogue and are representative of other episodes in the data corpus. Similarly, we interpreted directly from the transcript and memo [60] with the sensemaking episode about the conceptual inconsistency to document the inconsistency, sensemaking process, and resolution. The interpretative process was iterative as we engaged in multiple rounds of meetings and discussions among authors to establish internal consistency. As this episode comprises nearly the entire experimental time, we provide a rich narrative description of the group's sensemaking while highlighting the sensemaking details throughout the episode [61].

D. Focal cases description

The two selected cases in this study come from a single focal group of four students—Lisa, Chloe, Ethan, and Mika (pseudonyms)—engaging in the third of four lab investigations in the first-semester IPLS lab course. In this lab investigation, students are broadly tasked with developing an experiment that studies the Brownian motion of synthetic microspheres suspended in fluid and using this as a simple model of a chosen biological phenomenon. The focal group elected to study how the concentration of synthetic microspheres within a fluid impacts their Brownian motion; they chose to use this experiment as a model to investigate how medical drugs diffuse through the body. The students' initial hypothesis[4] was that higher concentrations of microspheres would have "greater Brownian motion." To explore this, the group collected videos of three different concentrations of suspended microspheres using a microscope with an attached video camera. They then used image tracking software to obtain the coordinate positions of the observed microspheres over time, then used spreadsheet programs to calculate the diffusion coefficients, a physical parameter indicative of a system's Brownian motion. Overall, the student group spent roughly 3.5 h engaging in formal experimentation across two lab weeks (1 h during week 1 and 2.5 h during week 2); the third week of lab time was designated for lab report peer review and any additional presentations, though no formal lab experimentation occurred.

VI. RESULTS

A. Focal case 1: Sensemaking about a conceptual inconsistency

The group starts their experiment by questioning whether different concentrations of microspheres in solution exhibit different rates of Brownian motion, resulting in an experimental design plan to investigate this topic empirically. This sensemaking episode begins when the group recognizes an inconsistency between their hypothesis and evidence while collecting videos of their microsphere samples using a microscope and attached video camera. In this case, the group's inconsistency arises when they observe high concentration microsphere samples moving more slowly than lower concentration microsphere samples, contradicting their initial hypothesis of a direct correlation between microsphere concentration and rate of Brownian motion. The sensemaking continues throughout the group's data collection and analysis, culminating after they finalize their experimental results and begin preparing their scientific arguments.

1. *Sensemaking initiated: Recognizing inconsistency and expanding hypothesis*

Roughly 38 min into week 1's formal data collection, the group recognizes an inconsistency between their observational evidence and initial hypothesis. Their observational evidence suggested that microspheres in a higher concentration sample were moving more slowly than those in a prior, lower concentration sample, while their initial hypothesis centered on higher microsphere concentration implying greater Brownian motion (Ethan: "We predict that higher concentrations, of microspheres will result in greater Brownian motion and diffusion in water."). Here, Lisa observes the inconsistency and describes it to her peers:

1036	Lisa	⟨*Looking at computer screen*⟩ No, it's moving, ⟨*briefly looks at group members*⟩ but not very much and we thought it would move more if there's more. I mean, maybe our hypothesis was wrong, but that < doesn't ma:ke sen:se>?

In the dialogue, Lisa shifts quickly between collecting experimental evidence by looking at the computer screen where she is observing the microsphere motion ("No, it's moving"), comparing the evidence with their initial hypothesis ("we thought it would move more if there's more"), and recognizing an inconsistency between the two ("maybe our hypothesis was wrong."). When positing that their hypothesis may be incorrect, Lisa scratches her head, speaks slower, and raises her vocal pitch, which we interpret as an additional corroboration that she recognizes an inconsistency between the group's observations and the initial hypothesis that she cannot immediately reconcile. Subsequently, the group continues with data collection through the remainder of week 1's experimentation without additional sensemaking about this conceptual inconsistency (much of the remaining week 1 experimental period consists of focal case 2, discussed in the Sec. VI B).

Early in week 2's experimentation, the group digs deeper into the inconsistency and expands their initial hypothesis with more detail, at times incorporating mechanistic elements. Roughly seven minutes into week 2 data collection, Mika asks what evidence one might see to support their

[4]The student group, like others in the data corpus, often used various terminology interchangeably when referring to their initial hypotheses or constructed explanations. Throughout the remainder of the paper, we utilize the term "hypothesis" to describe students' early proposed assumptions or predictions for their experimental results. Although students sometimes use the term hypothesis during dialogue, they also use the words "claim" and other terms. We use the term "explanation" to describe the product of students' efforts to build causal reasoning or scientific justification of their studied phenomenon [62]. Similarly, students infrequently use the term explanation during dialogue, instead utilizing terms such as claim or "reasoning."

hypothesis ("Can you explain what we want to see?"). Other students do not directly address Mika's question, but her question does prompt the group to add additional mechanistic elements to the explanation. Lisa suggests that the higher concentration microspheres will be "pinballing off of each other more," resulting in faster Brownian motion. Ethan adds that since the high concentration microspheres are "really highly concentrated, they are going to try to get somewhere else" and "they're in a really highly compacted area, and they don't want to be." He juxtaposes that idea with low concentration samples where "there's not a lot of spheres, … there's not much going on there." In these explanations for why their hypothesis might be correct, Lisa and Ethan recognize entities (microspheres) and activities (pinballing; faster motion) within a spatially organized structure (concentration), thereby suggesting mechanistic explanations [36]. Also, Ethan may be hinting toward an explanation that incorporates elements of concentration gradients and diffusion. Underlying this reasoning, perhaps Ethan does not distinguish between a concentration gradient within a single sample and multiple independent samples of varying concentrations. After this, the group transitions back to further data collection with no formal resolution.

To summarize above, initial sensemaking was rooted in an inconsistency in the students' understanding of a system or phenomena [21]. Similarly, other groups in our data corpus grappled with similar inconsistencies between their early experimental evidence and hypothesis, for instance, involving how the Brownian motion of microspheres was impacted by varying the fluid viscosity. More broadly, in lab courses such as this one, these inconsistencies invoke additional complexity since students are simultaneously grappling with tangible physical systems and underlying physical concepts; students may be developing hypotheses of their physical systems aligned with their conceptual understanding, which they then compare to their experimental evidence from the physical system [6]. Thus, the complexity of the setting and potential inconsistencies between hypotheses and evidence can spark opportunities for rich scientific sensemaking.

2. Sensemaking continued: Juxtaposing evidence and hypothesis

As the group's experimentation continues in week 2, they iterate between collecting evidence, acknowledging the recurring inconsistency between ongoing evidence and hypothesis, and attempting to resolve it by expanding their hypothesis to align with and explain their evidence. For example, after gathering additional evidence for roughly 32 minutes, the group is again presented with the inconsistency that the microspheres are not moving as quickly as expected (Chloe: "@@ They—they're not mov:ing." Lisa: "They're moving a litt:le.") Lisa responds by suggesting a new explanation that modifies, but does not replace, their existing hypothesis to align with this additional evidence:

1860 Lisa I wonder if that's like—is there a point where
like—and this could be a different >part of
the hypothesis, part of our claim<, is that
there's a point where it becomes too*ooo*
much of a dosage, you know, where
it ceases to be helpful because it's like,
not diffusing well. Is that a possibility?
I don't know. I'm just throwing that out
there.

Lisa suggests that highly dense concentrations may limit microsphere motion and speed since microspheres cannot move directionally without colliding with other microspheres. However, this explanation is not explored further, evident by Ethan's quick deference to the quantitative data analysis he anticipates later in their experimentation ("We'll find out when you get the numbers back.").

As Chloe and Lisa continue collecting additional evidence from their microsphere videos, Mika and Ethan begin assessing whether there are erroneous elements of their hypothesis since their more detailed hypothesis still cannot explain their ongoing evidence. While reviewing external resources (e.g., reading an unknown website on their phones) to investigate the mechanisms associated with Brownian motion and diffusion, Mika and Ethan come to an agreement that the reasoning for their hypothesis—in which they correlated concentration gradients within a single sample to varying concentrations in distinct samples—is incorrect since this reasoning assumes a concentration gradient in each individual sample:

1926 Ethan The—the—the issue I think is (2.5), the issue
I think is that they're all—in each video,
they're all equally distributed throughout.
So, in the body, in the body, when you take
like a—some type of medication, uh, uh,
whatever is in that pill or whatever, is
flowing through your body and it's
not equally distributed throughout your body.
So it's gonna keep on going, because
there's less of it in front of it.
If that makes sense.

Here, Ethan recognizes that each observed sample has microspheres "equally distributed throughout," contradicting their earlier reasoning that microspheres moved via concentration gradients. Subsequently, Ethan and Mika reject their hypothesis and plan to use their experimental evidence in later sensemaking to provide a new explanation:

1959	Ethan	I think that's the issue that we're gonna run into. Because if they're already equally distributed, if it's in equilibrium, then they're all gonna go pretty slow. So I guess wh—I guess what we're really looking at is just (5.0) I guess, the mor—the le—the more amount of spheres, I guess, they're gonna—are gonna (2.0) move less? I don't know. That's why, I—I think—you know what, we'll get the data. I think we find out, and then I think we're gonna have to make a new claim @@@ that probably has little to do with what we were originally looking at, you know?

Ethan hesitantly offers a new hypothesis that "the more amount of spheres … are gonna move less?" Ethan's comments suggest that he is likely uncertain about how to reconcile their initial hypothesis with ongoing evidence since the reasoning supporting their hypothesis is no longer accurate. Possibly because he is uncertain about alternative explanations, he defers further sensemaking until they complete quantitative data analysis and suggests that their final experimental results (the "new claim") will significantly differ from their initial hypothesis. As they return to data collection and analysis, the group continues to critique their original hypothesis, though they offer no new explanations that could explain their evidence suggesting microsphere concentration and rate of Brownian motion are inversely proportional. This gap in explanation building related to the conceptual inconsistency persists for roughly 90 min until they complete their data analysis and obtain final experimental results, which show an inverse relationship between microsphere concentration and rate of Brownian motion, contradicting their initial hypothesis.

During the sensemaking process described above, the students construct and critique their hypothesis while exploring their experimental evidence. The group's hypothesis construction and critique shift between developing a more detailed hypothesis, including Lisa's proposed modification (line 1860), recognizing ongoing issues with their more detailed hypothesis, and ultimately rejecting this hypothesis altogether since it continues to not align with their ongoing experimental evidence. As a result, the group comes to an agreement that a new explanation is necessary to explain their results, and in this case, they wait until obtaining conclusive experimental results to again take up constructing such an explanation. Similar to how students in this group are building up their hypothesis by "shopping for ideas" from external resources (i.e., websites; course notes) as they integrate new information into an existing understanding [34], other groups in the data set embarked on a similar process. For instance, we saw groups review peer-reviewed articles to explain their experimental results and expanded hypotheses, all in service of resolving previously encountered inconsistencies. In these ways, sensemaking often involves an iterative process of constructing and critiquing hypotheses while simultaneously accounting for new evidence and new information to resolve an inconsistency in one's understanding [21,35].

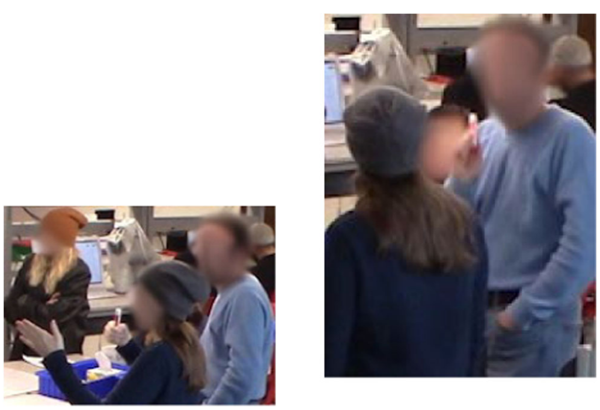
3. Sensemaking resolved: Achieving resolution through explanation building

Having obtained final results midway through week 2 (roughly 3 h into formal experimentation), the group begins working to build a new explanation that matches their evidence: high concentrations of microspheres have low rates of Brownian motion, and vice versa (Ethan: "… we're trying to justify our results for Brown—because they're flip flopped from what we thought they should be."). Lisa connects her earlier explanation that crowded microspheres have limited area for motion with Ethan's explanation that their samples have constant concentrations without gradients (Lisa: "Because you're assuming diffusion if the spheres are moving into an area where there's no longer space. But they were all crowded together, so there wasn't really much … area for them to move.") Shortly after, the LA joins the group's discussion and prompts them to describe the mechanisms associated with the microsphere motion (LA: "What makes them erratically bounce either way in the water?"), initiating a detailed conversation where the group builds a new mechanistic explanation for their results. The group acknowledges for the first time that the microspheres are colliding with fluid particles, in addition to other microspheres (Lisa: "Well it's running into either, like, fluid that it's in or other spheres around it.") Later, Lisa compares the highly concentrated microspheres to "like when you're in a crowded elevator, and you can't move very much." By providing an analogy, Lisa introduces a mechanism to which the group can compare as they continue building a new explanation of their results. Ethan continues to explore this line of reasoning:

3157	Ethan	So in this one ⟨the low concentration sample⟩, there's a lot more fluid particles, because there's—there's way more space between the microspheres, right?
3158	Lisa	M'kay. I buy that.
3159	Ethan	So, so then, because there's more of the fluid particles, they're able to hit those microspheres, there's a lot more to hit these microspheres=
3160	LA	=That's true=
3161	Ethan	=and move them around, right? So that would, say, that would support the reasoning … that there's more, Brownian motion.

At this point, the group has developed a plausible explanation that supports their results and their experiment's underlying mechanisms: lower concentration microsphere samples have more fluid particles surrounding them, which "hit those microspheres" and cause greater rates of Brownian motion. However, Ethan pushes beyond this reasoning, asking why their explanation does not account for microsphere interactions. The group responds with additional explanation building that involves animated, embodied action (see Fig. 1):

3166	Lisa	Maybe because they're more dense, so it changes the motion more? Because, like, when it bumps into a fluid particle, it like, it's not gonna—it's just gonna change its course ⟨*slides hands past each other*⟩ versus like stop it from moving. Whereas it runs ⟨*punches fist into open hand*⟩ into something with its same size, it's kinda gonna stop ( ) in its tracks =
3167	LA	= she's got a good point, since I'm comfortable with Lisa, I don't mind doing this, ⟨*picks up marker*⟩ here's the fluid particle ⟨*tosses marker at Lisa's shoulder*⟩ did you move that way, just a little bit?
3168	Lisa	I mean maybe.
3169	LA	How about now? ⟨*Bumps shoulders with Lisa*⟩
	Lisa	Yeah.
3170	LA	So tha—because they're more massive (2.0) a bigger particle running into you will have a bigger ( ) than a smaller—



(a) (b)

FIG. 1. Group's embodied action during explanation building. (a) Lisa punching fist into open hand, representing a microsphere-microsphere collision (see line 3166) and (b) LA throwing a marker at Lisa, representing a fluid particle running into a microsphere (see line 3167).

In line 3166, Lisa provides an explanation for Ethan's inquiry while also adding further mechanistic detail to her earlier statement of microspheres being "crowded together, so there wasn't really much … area for them to move." Here, Lisa recognizes that microsphere collisions with other microspheres will limit the microsphere's motion much more than when the microsphere collides with a fluid particle. Here, we hypothesize that Lisa is incorporating her intuitive understanding of conservation of momentum and collisions to support her reasoning; she may also be implicitly correlating the mean-free path of the microspheres with the rate of Brownian motion. In lines 3167–3170, the LA supports Lisa's reasoning by using embodied action to model the discussed mechanisms. First, in line 3167, the LA tosses a marker at Lisa, causing almost no reaction from Lisa. Then, the TA bumps shoulders with Lisa, causing her to shift her footing to keep balanced. In this embodied action, the LA models microsphere-fluid particle motion with the marker-body collision and models the microsphere-microsphere collision with the body-body collision. We argue that this embodied action supported the students' reasoning, in line with existing literature [38–40], as the group expresses contentment and relief with this final explanation building, evident by their body language, tone, and comments ("What you said makes so much more sense" and "So I can explain that in our whiteboard, 'cause that makes sense."). Subsequently, they transition to preparing their final whiteboard presentations using their new explanation and reasoning. We interpret that the group views their sensemaking as successful since they "figured out" a resolution for their inconsistency by building a new explanation for their results that incorporates multiple elements of mechanistic reasoning.

In this final piece of the sensemaking process, the group achieved resolution of the inconsistency by building an explanation that incorporated mechanistic elements. Their final explanation relied on their efforts to make sense of mechanisms associated with Brownian motion by building analogies and exploring the entities and activities involved in the system. Despite the group viewing their experimental evidence and results as accurately representing the physical system (i.e., there were no experimental errors), they do not assess their data collection or analytical procedures for potential errors that could lead to incorrect empirical results. Yet, importantly, by the end, students have resolved the inconsistency and developed an explanation, thereby concluding the sensemaking process in which they aimed to "figure something out" [21]. For this group, building an explanation involves numerous "moves" documented in existing PER literature, including mechanistic reasoning (e.g., identifying the properties and organization of entities and activities [36]), incorporating formal or common-sense knowledge, and appealing to authority [20], many of which were also utilized by other groups. For example, we saw other groups incorporate knowledge from various sources into their sensemaking. Across all groups, common to the sensemaking was an eventual resolution of the inconsistencies and final explanations that were supported by the experimental results.

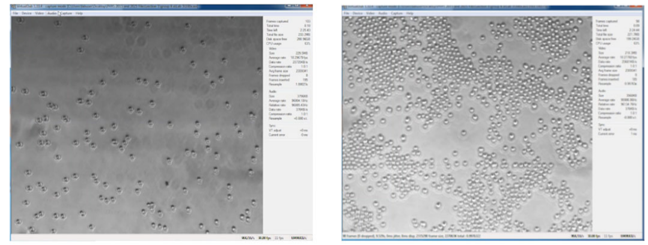
4. Sensemaking about conceptual inconsistencies across the data corpus

We saw many groups encounter inconsistencies between their hypotheses and evidence. While the focal group presented here identified an inconsistency early in data collection, other groups recognized an inconsistency after obtaining final experimental results that expressly contradicted their hypothesis. Still other groups did not recognize an explicit inconsistency between hypothesis and evidence, yet the comparison between the two eventually led to explanation building in efforts to align their hypothesis more closely with ongoing evidence and final results. We also observed that a small number of groups were especially resistant to rejecting their initial hypotheses in favor of new explanations, even when their results were inconclusive or directly contradictory.

When groups began engaging in sensemaking, they often did so by reinforcing their hypotheses with additional conceptual information, including from external resources (e.g., websites, peer-reviewed journal articles). In some instances, this reinforcement led to successful resolution through further explanation building, while in others, like the focal group above, the hypothesis is ultimately rejected in favor of new alternative explanations. Finally, we saw groups consistently work towards building explanations to explain their results, possibly in response to explicit course expectations to articulate the scientific reasoning behind one's results. While not exclusively, explanation building sometimes happened as groups finalized their experimental results and prepared their scientific arguments to present to their peers and instructors. Thus, we recognize that a range of conceptual sensemaking contributed to richer explanations across the data corpus, with sometimes initial hypotheses being rejected and sometimes refined. Furthermore, across all the data, many groups' sensemaking episodes were productive based on the students expressing confidence and satisfaction in their explanation and resolution [20]. For our focal group, similar to others in the data corpus, their sensemaking about a conceptual inconsistency was productive in that their sensemaking prompted additional experimentation and explanation building until they ultimately built a detailed mechanistic explanation of their studied phenomenon. Overall, we see that sensemaking about conceptual inconsistencies was common across many groups, across different labs within the course, and at various points in the experimental process, including collecting evidence, analyzing data, and exploring results.

B. Focal case 2: Sensemaking about procedural inconsistencies

As a comparison, this case highlights how the focal group engaged in sensemaking about a procedural inconsistency early in their experimentation as they collected videos of their microsphere samples. This episode, representative of similar episodes of sensemaking about procedural inconsistencies in the data corpus, involved the students recognizing a procedural inconsistency and then iterating between generating potential causes and proposing, carrying out, and assessing procedures to test the potential causes. This form of sensemaking is largely synonymous with troubleshooting, explored in the PER literature [16]. This episode highlights the complexity of groups' sensemaking about procedural inconsistencies, which in this case involves multiple inconsistencies occurring simultaneously. Here, we document the focal group's successes and struggles while engaging in sensemaking to navigate and resolve multiple related procedural inconsistencies. This sensemaking episode occurred during roughly the last 30 min of week 1 experimentation time.



(a) (b)

FIG. 2. Screen captures of group's 100× sample, from which Lisa begins recognizing a procedural inconsistency. (a) Screen capture of group's 100× sample shortly after pipetting and (b) screen capture of group's 100× sample after several minutes.

1. Sensemaking initiated: Recognizing an inconsistency

The students first recognize an inconsistency in the concentration of microspheres as they collect videos of microsphere samples exhibiting Brownian motion. During data collection, Lisa notices that the concentration of microspheres in the 100× solution has increased significantly since they pipetted it into the microslide[5] (see Fig. 2; Lisa: "Okay, this looks different than it did last time. They've, like, coalesced more, >look at that<. Wow. (2.5) We're looking at the 100, right? They have coalesced a lot more than the last time.") At first, the other group members were less attentive to this "concentration change over time" inconsistency, and Lisa again points at the computer screen and describes the inconsistency that the concentration in the fluid sample is higher than when they collected their initial trial video with the same sample (Lisa: "—that's interesting. They've coalesced a ton <in only like> two minutes

[5]In this instance, students used a microscope well slide, or a microslide which has eight individual walled chambers to house fluid samples. These types of microslides were commonly utilized in groups' lab investigations.

FIG. 3. Lisa expresses frustration and confusion while recognizing procedural inconsistency.

since @we @took the last video.") As her peers are still focused elsewhere, another inconsistency arises. The group collectively recognizes that the 1000× concentration sample has significantly fewer microspheres than the 100× and 500× samples, contrary to their expectations (Lisa: "Ugh *hhhh*, 'cause there's not, like, very many in that shot"). While recognizing this "concentration trend" inconsistency, Lisa raises the tone of her voice, furrows her brow, and grabs her hat with both hands, which suggests frustration or confusion (see Fig. 3). In response, her group peers pay attention by leaning toward the computer screen and looking at the sample, suggesting they are now aware of the inconsistency. Though Lisa and the group initially encountered two inconsistencies—the former "concentration change over time" inconsistency and the latter "concentration trend" inconsistency—the group's subsequent sensemaking focuses on the resolution of only the second identified inconsistency, that the 1000× microsphere sample has fewer microspheres than the other samples, while the group expected the 1000× sample to have the highest microsphere concentration.

In this initial sensemaking, the process commences with students recognizing an inconsistency with their experimental procedures or apparatus [16]. Sensemaking about a procedural inconsistency is common in lab settings, particularly in environments where students use sophisticated apparatus and are responsible for developing and carrying out their own procedures. We noticed other groups encounter similar inconsistencies, such as involving "jumps" in videos involving position tracking of objects. Furthermore, comparing this instance of sensemaking about a procedural inconsistency with the conceptual inconsistency discussed previously, we note that this inconsistency was quickly and clearly recognized compared to the prior case where the student struggled to fully recognize and diagnose the conceptual inconsistency between hypothesis and evidence. Importantly, for these instances of sensemaking about a procedural inconsistency, recognizing inconsistencies in one's experiment is a crucial practical skill that leads to proficient and successful experimentation [16,41].

2. Sensemaking continued: Proposing and testing potential causes

Given their recognition of the concentration trend inconsistency, the group next proposes potential causes of the 1000× microsphere sample having fewer microspheres than other samples. Lisa suggests that she may have shaken the solution vial too much before pipetting it into the chambered microslide (Lisa: "Did I shake it too much?"). Chloe asks if the sample was shaken at all (Chloe: "Was this even shaken? This was shaken, right?"), then proposes that the microspheres in the sample vial may have settled before pipetting, potentially resulting in a lower concentration in their observed sample (Chloe: "It might have just settled.") At this point, the LA passes by the group, and Lisa asks the LA, "This is supposed to be a higher concentration… Did I shake it too much?" In response, the LA begins proposing other potential causes for the inconsistency, including the possibility that the group unintentionally used an incorrect concentration, microsphere size ("Same size spheres?"), magnification ("You're on 40 times magnification?"), fluid viscosity ("So the viscosity's the same—"), or mixing method ("They both mixed well?"). As the LA poses these successive possibilities, the group affirms that their experimental procedure accounted for many of them, and they quickly move on to other potential causes. Given the potential causes for the inconsistency, the group tests them by iteratively proposing, carrying out, and assessing various procedures. First, given the potential cause of viewing a low-concentration area of a sample with varying concentration (LA: "Maybe you just have <a> sparse section"), the LA proposes that the group "cruise around" the microscope viewing area to observe different sections in their microsphere sample. This prompts the group to adjust the microscope dials to observe different areas in the sample. Quickly, the group assesses that the sample's microspheres are equally distributed throughout, eliminating this potential cause and driving further testing. To test the potential causes related to inconsistent solution mixing, Ethan proposes a new procedure to use a constant vortex time for all microsphere samples to maintain consistency (Ethan: "So maybe—do we need like, a constant, vortex time? Should we have done, like, twenty seconds of vortex for each of them?") However, before the group further pursues this proposed procedure, the LA refutes it, saying "If they're sufficiently mixed, they're sufficiently mixed," and the group interprets that mixing alone may not be a viable cause for their inconsistency.

Shortly after, the group begins iterating between proposing potential causes and testing these causes in short succession. The LA asks, "Is there a higher concentration at a different focal plane, maybe it was overlooked?," suggesting that the inconsistency may be caused by the samples having different microsphere concentrations at various depths. He then proposes a new testing procedure: “Have you checked to—say your fluid’s this deep, you looked in all the different planes, lest—are you j—I don’t know, are you near the bottom where they’re kind of rubbing?” The group does not implement this testing procedure, but it leads to the group proposing related potential causes. Ethan posits that different volumes of fluid may cause the inconsistency in each sample: “what about, um, like, the <amount of fluid that we—that we put into each of them…>.” The group immediately tests this potential cause. Lisa picks up the chamber well and looks at it horizontally to compare the fluid levels; observing the fluid levels and assessing all prepared samples to be “pretty constant” in depth, the group dismisses this cause. Finally, with a continued focus on the fluid level as a potential cause, the LA proposes adding more fluid to the samples to increase the total concentration: “Why don’t you *uuuu* (0.5) grab some more from it? Maybe you just got a (0.4) sparser section? Or—eh—bum—maybe it’s just more concentrated, eh, somewhe:re else:.” The group implements this suggestion, with Chloe adding “<twoooo things>” of an unknown volume to the existing 1000× microsphere sample. Assessing this new procedure, the group observes that the sample still displays an unexpected microsphere concentration, meaning they have not identified the cause for their inconsistency. At this point, the group (and LA) express frustration over the unresolved inconsistency (Chloe: “I just want to redo it completely.”), likely exacerbated by the limited remaining class time and continued unsuccessful sensemaking.

In this stage of the sensemaking process, we see students working to determine the cause of their recognized inconsistency. We know that resolving an inconsistency in one’s procedure or apparatus necessitates determining how and why the inconsistency came to be [41,42], which was seen in the data above. The group iteratively proposes, carries out, and assesses diagnostic procedures to test multiple potential causes, synonymous with elements of experimental troubleshooting [41]. More generally, proposing and iteratively testing potential causes of a procedural inconsistency was a common feature throughout the data corpus, as students encountered a variety of inconsistencies with the software that involved, for instance, light contrast issues and miscalculated software parameters. While many groups successfully diagnose the causes of their inconsistencies, the focal group here was ultimately unsuccessful even after multiple rounds of proposing and testing potential causes with LA support. Furthermore, compared to the previously discussed case of sensemaking about a conceptual inconsistency where digging into the inconsistency resulted in an expanded hypothesis and explanation, the sensemaking case here remained focused on the data collection procedures without focusing on potential underlying mechanisms associated with the inconsistency.

3. Sensemaking resolved: An incomplete resolution

The group ultimately struggles to produce a complete resolution with an identified cause and proper revision or repair, possibly in large part due to their challenges in identifying a cause for the inconsistency. As the LA and group summarize the inconsistency to the TA (“They went from 100 concentration to 1000. We’re seeing way fewer microspheres.”), the TA immediately recognizes that the group was interpreting the microsphere labels (e.g., 1000×) as microsphere concentration (microspheres per volume) instead of the fluid’s dilution factor (volume per sphere): “It’s dilu:tion.” In response, the group members engage in sensemaking about this new information (Ethan: “…so the higher the <concentration the more>, wate:r.” and Lisa: “Oh I thou:ght it was the conc—I thought there was more <spheres> in the one thous:and—.”) and agree that this new information resolves their inconsistency (Ethan: “That answers the questi:on.”) However, from the researchers’ perspective, the TA may have identified a new cause for the group’s inconsistency, which the group interpreted as a formal resolution. In order to provide a more comprehensive resolution of this inconsistency, the group could have tested this potential cause by preparing and observing new samples of each labeling to verify the TA’s new information. However, for any number of reasons, including likely limited remaining time, the group counts the inconsistency as resolved.

The group’s struggles to identify a cause and obtain full resolution may also be due to focusing on one inconsistency while skipping over another. Recall that early on in this sensemaking process, Lisa recognized two related inconsistencies, the concentration change over time inconsistency and the concentration trend inconsistency, though the group only worked to resolve the latter inconsistency of the 1000× sample having a lower concentration than other samples. Just as they are expressing the relief of achieving resolution through TA consultation, the group again encounters the omitted inconsistency while observing a newly prepared 1000× sample:

1224	Lisa	So wait, which one’:s th:at th:en? ⟨*referring to the new sample just identified on microscope*⟩ That one’s one thousa:nd?
1225	Ethan	Wai—wha–=
1226	Lisa	<Wait> a seco:nd. Then how come there’s mo:re in this o:ne n:ow. Ah*h:hhhhhh*=
1227	Ethan	=What is this—what is this one, that=
1228	Lisa	This is the one thousa:nd, right? It’s the one we just used. °I’m so confused.° (2.0)
1229	Chloe	Because (our) 1000 looked very sparse.
1230	Lisa	<Yea:h>, but now the one thousand looks super, no:t spa:rse.

Here, Lisa and Ethan recognize that the newly prepared 1000× sample has a significantly higher concentration than earlier prepared 1000× samples; in essence, we interpret that the group is encountering a variation of the concentration change over time inconsistency of individual microsphere samples having different concentrations over time. Lisa's tone and questioning suggest that she is surprised by her observations of the sample's concentration, likely because she expected the TA's new information to have resolved issues with microsphere concentrations. With class time coming to a close, the group did not deliberate on this recurring inconsistency and how it may be related to the TA's new information. However, we see evidence that this unresolved sensemaking process was productive in that it produced new experimental procedures; the group began the following week's experimental time by discussing new procedures to incorporate consistent vortexing times and pipetting methods while preparing microsphere samples. One interpretation is that, while the group was unable to identify, test, and resolve the cause for their inconsistencies, they do recognize a need for more detailed and intentional experimental procedures to mitigate future procedural inconsistencies.

In this final piece of the sensemaking process, we see contradictory interpretations from the group of the success of their procedural sensemaking. On the one hand, the group views the TA's new information as fully resolving their inconsistency, but on the other hand, they encounter a recurring inconsistency that frustrates and perplexes them. One interpretation of the challenges this group faced in their sensemaking process was that their primary focus on a single inconsistency may have clouded a full resolution to the complex network of multiple inconsistencies the group faced during this phase of experimentation. It could be that the group did not properly identify the cause for their inconsistency because they were not explicit in identifying and describing the full slate of inconsistencies they were observing. Additionally, the group's recognition of the inconsistency as procedural may have clouded attempts to make sense of underlying physical concepts or mechanisms toward achieving a more conceptually rooted resolution. Regardless of the reasons for their unresolved experimental inconsistency, we do see evidence that their sensemaking efforts were not altogether unproductive. When picking up in their second week of lab time, the group spends time discussing more detailed and consistent experimental procedures to carry out, which may stem from the challenges they faced with their earlier sensemaking.

4. Sensemaking about procedural inconsistencies across the data corpus

The essence of this group's sensemaking about procedural inconsistencies was largely analogous to many other groups within the data corpus. Other groups also recognized inconsistencies related to their video collection procedures, as well as computational methods that were used in their analysis. We also noticed some groups' inconsistencies involved procedures to generate graphical representations and were related to the inadvertent inclusion of erroneous data. These other groups also often iteratively shifted between proposing potential causes for their procedural inconsistencies and then testing those causes. There was a range of approaches to resolving these inconsistencies, with the inconsistency's cause sometimes being self-evident and thus requiring no exploration. Other times students skipped determining the cause of an inconsistency and focused on repairing and revising their procedures or apparatus. Thus, sometimes the attempts at resolution were explicit and well documented, while other times, it was more implicit with minimal discussion. Across the data corpus, there were various ways in which students' sensemaking about procedural inconsistencies was productive, either via resolution to the inconsistency, greater attention to experimental procedures and apparatus, or possibly building knowledge procedural or system knowledge [16].

Finally, we also noticed some instances of inconsistencies that shared features of both procedural and conceptual inconsistencies. In these instances, there tended to be an inconsistency between a hypothesis and evidence, similar to the conceptual inconsistencies, but yet it occurred over a shorter timescale and was resolved quickly with little explanation building, more similar to a procedural inconsistency. Similarly, groups sometimes encountered inconsistencies that they approached as procedural, resulting in efforts to produce procedural resolutions even if there may have been underlying physical mechanisms factoring into the recognized inconsistency. Thus, we recognize that the two types of inconsistencies found in our data are likely not opposites but instead occur along a continuum in which there is a range of ways students engage in sensemaking about an inconsistency.

VII. DISCUSSION

This study uses a sensemaking framework to identify and characterize what forms of inconsistencies *about* which students engage in sensemaking during reform-based introductory physics lab experimentation, and what *moves* students enact during their sensemaking to achieve resolution. In doing so, this study helps fills a gap in existing literature related to students' moment-by-moment sensemaking processes in reform-based introductory physics laboratory courses. Existing research has primarily focused on two ends of an empirical spectrum: some studies focus astutely on student reasoning within single experimental processes, such as experimental design or troubleshooting, with limited attention to other parallel processes; other studies focus on characterizing students' broad experimental processes collectively using comprehensive frameworks. This study is situated squarely between these two termini, focusing on students' moment-by-moment sensemaking as they progress through multiple experimental processes while conducting self-directed experiments. Broadly, this work pushes forward our understanding of student reasoning in introductory physics courses by providing detailed analysis of what initiates student sensemaking (inconsistencies) and what students do (enacting various sensemaking moves) while engaged in sensemaking processes. Thereby, these results paint a more authentic picture of the inherent complexity of student sensemaking in this setting.

In this study, we documented students' recognition of inconsistencies as either conceptual or procedural in nature—in large part based on how students themselves described and responded to the inconsistencies—and we described the sensemaking moves students enacted to resolve each form of inconsistency. In focal case 1, the group's inconsistency was characterized as conceptual due to the group's comparisons of early experimental evidence with their initial hypothesis; this comparison led the group to determine that elements of their conceptual explanations within their initial hypothesis did not sufficiently explain their ongoing experimental evidence. Recognizing this inconsistency led the group to enact multiple sensemaking moves, including expanding their hypothesis while incorporating elements of mechanistic reasoning, juxtaposing evidence and hypothesis, critiquing and rejecting their hypothesis, and engaging in explanation building through mechanistic analogies and embodied action. By enacting these sensemaking moves, the group was able to resolve their conceptual inconsistency and build a new explanation of their experimental results, describing the mechanisms at play within a complex system of numerous microspheres exhibiting Brownian motion.

In comparison, focal case 2 presents the group sensemaking in response to a procedural inconsistency. This inconsistency was characterized as procedural due to the group's focus on their preparation and use of experimental apparatus and samples while recognizing and responding to the inconsistency in question. The group made sensemaking progress about the procedural inconsistency through sensemaking moves such as proposing potential causes for the inconsistency and iteratively testing these potential causes towards adjusting their apparatus or revising their procedures. Possibly due to the group's struggles recognizing and categorizing their inconsistency, the group was unable to complete the sensemaking process towards a complete resolution, though their sensemaking efforts did lead to more attentive procedural efforts in subsequent experimentation.

A primary takeaway from these results is that students enact a variety of sensemaking moves while sensemaking to resolve various forms of inconsistencies in reform-based physics lab courses. Possibly, these moves may function as underlying learning mechanisms causing forward momentum of the sensemaking process. If we consider these moves to be learning mechanisms, future work would benefit from exploring the impacts of these different moves on student learning processes and outcomes. As well, in both focal cases, recognizing and characterizing the inconsistencies was crucial in the subsequent sensemaking, yet there were differences in what the inconsistencies were *about* and *how* they were resolved. These cases exist within a larger data corpus where this type of sensemaking about inconsistencies was relatively common. Throughout the data corpus, we frequently observed similar patterns of students' recognition and categorization of inconsistencies and subsequent enactment of sensemaking moves towards achieving resolution. However, there were differences among groups' sensemaking episodes; sometimes the inconsistencies were resolved with ease, and sometimes it was a deeper struggle and not resolved immediately or at all.

A. Situating results within physics education research

The current study highlights the moment-by-moment progression of student sensemaking in reform-based lab courses, an understudied learning environment that is becoming increasingly prevalent as more institutions implement inquiry-based lab environments. This study goes beyond existing instructional labs PER literature by focusing on the inconsistencies that students encounter during laboratory experimentation and the subsequent sensemaking moves enacted to achieve resolution. Much of the existing labs-based PER literature focuses either on students' reasoning within single experimental processes, such as experimental design (e.g., [3,63]) or on the broad characteristics of student reasoning throughout an entire experiment (e.g., [6,19,64]). By elucidating what students are sensemaking about (conceptual and procedural inconsistencies) and how they go about this sensemaking (various sensemaking moves), this study builds new knowledge of student reasoning in physics instructional laboratory experimentation.

This study adds to the experimental modeling framework's descriptions of revisions by providing detailed moment-by-moment analyses of students' revision processes through sensemaking moves [6,18]. Using a sensemaking framework, we documented several sensemaking moves students enacted as they worked to resolve both procedural and conceptual inconsistencies. Efforts to resolve these two forms of inconsistencies closely align with the modeling framework's revision process, whereby students work to revise either the measurement model (procedural) or physical system model (conceptual) due to disagreement (inconsistencies) between the models. Specifically, our results from the procedural inconsistency case are similar to those produced by overlaying the Modeling Framework atop students' efforts to engage in experimental troubleshooting [16]. In both the literature on troubleshooting [16,41,42] and our data, students identify a problem (recognizing an inconsistency), propose potential causes, and test these causes to work towards repairing or revising procedures or equipment. However, our work sheds light on the challenges that can occur during procedural sensemaking and troubleshooting; while existing literature largely documents successful instances of experimental troubleshooting, focal case 2 highlights how difficulties recognizing and characterizing a single procedural inconsistency while encountering multiple inconsistencies simultaneously may result in unresolved sensemaking or incomplete troubleshooting. More broadly, this points to the possibility that students likely encounter multiple (sometimes related) inconsistencies simultaneously, rather than sequentially, making their sensemaking efforts considerably more complex. Similarly, our results from focal case 1 align closely with how the modeling framework describes students' efforts to revise their physical system models in light of disagreement between measurement and physical system models. Here, students worked to revise their initial hypotheses (physical system models) to more closely align with their ongoing experimental evidence (outcomes of measurement model). Importantly, this current study provides a more detailed moment-by-moment perspective of students' efforts to revise their physical system models by documenting both what sparks the revision process (conceptual inconsistency) and what actions students take to revise their models (various sensemaking moves).

We note that in both our study and the modeling framework, students have the important task of implicitly determining whether the inconsistency in question should be resolved through procedural or conceptual means. As we discussed in Sec. III, experimental inconsistencies likely lie on a spectrum ranging from more procedural to more conceptual [15,26–32], and where students implicitly situate their inconsistencies on this spectrum likely has a significant impact on the sensemaking processes they enact to resolve them. Empirically, we argue that it is crucial that PER researchers focus astutely on identifying the factors that influence students' categorization of inconsistencies and their resultant sensemaking processes. Likewise, we propose that laboratory instructors' pedagogy may benefit from increased focus and support for students as they recognize, categorize, and then resolve inconsistencies during experimentation.

From a different perspective, our results also align with and extend the empirical landscape of Odden and Russ's sensemaking framework [9,20] by demonstrating that their prior approach to student sensemaking in more controlled learning environments applies to these physics lab courses where students experienced experimental and intellectual agency. Also, importantly, Odden and Russ [20,21] identified various parameters relevant to sensemaking that were seen in our data, namely, entry conditions (inconsistencies), moves (expanding and critiquing hypotheses, building analogies, embodied action), and exit conditions (mechanistic explanation, approval from authority). Future work would benefit from exploring the utility of the sensemaking framework in other learning environments (e.g., informal physics settings, upper-division lab courses) and to what extent different sensemaking moves are more or less application to these varied learning environments.

The focal cases explored here involved students engaged in sensemaking related to Brownian motion, a pervasive phenomenon across multiple disciplines, including physics, biology, chemistry, and mathematics, but understudied in the PER literature. While educational researchers from other STEM disciplines have explored how students' conceptual knowledge of Brownian motion transfers to other contexts [65,66] or how computational simulations can be used to illustrate Brownian motion [67,68], few studies have explored students' reasoning processes about Brownian motion at a mechanistic level in experimental or practical contexts. Our data analysis captured some of the complexities in students' reasoning about this important phenomenon, and future work may pave the way for highlighting reasoning across disciplines.

B. Limitations and future directions

Given the study's emphasis and methods, various important pieces of the classroom setting were minimized in the analysis. In the second case, TAs and LAs had an impact on the recognition of the inconsistencies and subsequent sensemaking processes but were not the focus of the analysis. Other factors also impacted the process, including the group social dynamics, the arrangement of equipment and materials in the room, individual group members' identities, and students' experiences and background knowledge, all of which might be important directions for future research. Finally, we note that the sensemaking about the second focal case's procedural inconsistency might contain an underexplored conceptual inconsistency as the students focused on modifying experimental procedures and omitted discussion about underlying physical mechanisms. This further reinforces the potential spectrum of forms of inconsistencies rather than procedural and conceptual inconsistencies existing dichotomously; future work may benefit from further exploration into the relationship between these two types of sensemaking.

Moving forward, this study's results raise questions for future research in physics lab contexts. What other forms of inconsistencies might students encounter during physics lab experimentation? With a multiplicity of types of physics lab courses, each with distinct learning goals and structures, we suspect that there are likely many other inconsistencies that students encounter that may prompt sensemaking. What cues might prompt students to recognize experimental inconsistencies in lab contexts and take up sensemaking to resolve them? What other sensemaking moves might students enact when resolving inconsistencies in physics lab courses? We identified several sensemaking moves in this study, but there are undoubtedly additional moves students may choose to enact in response to recognized inconsistencies, especially with forms of inconsistencies not documented here.

There are also several pedagogical questions stemming from this work. How can instructors recognize inconsistencies and sensemaking moves as they occur in student experimentation? If instructors hope to engage their students in sensemaking during experimentation, it is crucial that they have sufficient knowledge and strategies for helping students recognize and respond to experimental inconsistencies through sensemaking. Also, how might group dynamics influence student sensemaking? Do students' group roles, whether explicitly defined or implicitly self-assumed, impact who has ownership over the sensemaking space? When recognizing students are struggling with an experimental inconsistency, how can laboratory instructors prompt and support student sensemaking that works towards resolution? While it was not the analytical focus, focal case 1 documented how the LA supported students as they engaged in sensemaking through mechanistic reasoning and explanation building. Future work could more carefully document successful strategies instructors enact to elicit similar forms of sensemaking. Finally, what degree of experimental agency is optimal for providing students opportunities to engage in rich sensemaking in introductory physics lab courses while maintaining appropriate pedagogical scaffolding and expectations? From this study's results and the questions they raise, we see that documenting these forms of student sensemaking in a reform-based lab course contributes to a larger dialogue about the nature and goals of physics lab instruction as we aim to provide students with authentic physics learning environments.

ACKNOWLEDGMENTS

This material is based upon work supported by National Science Foundation (NSF) Grant No. 1938721 and National Science Foundation Graduate Research Fellowship Program (NSF-GRFP) Grant No. 1747505. 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. We thank Tamara Young for providing insightful feedback that greatly improved this manuscript. We also thank Molly Griston for contributing to the analysis and coding for this project. We thank David P. Goldenberg and members of our Advisory Board—Leema K. Berland, Melanie M. Cooper, Benjamin D. Geller, Laurie E. McNeil, Vashti Sawtelle, and Bethany R. Wilcox—for providing invaluable feedback about many elements of this study. We also thank Adam J. Beehler, Kevin Davenport, David Thomas, Andy Lee, and the outstanding teaching and learning assistants involved in the IPLS course preparations and instruction. We would finally like to thank the student participants who provided consent to participate in this research study.

[1] AAPT Committee on Laboratories, *AAPT Recommendations for the Undergraduate Physics Laboratory Curriculum* (American Association of Physics Teachers, College Park, MD, 2014).

[2] N. Holmes and H. Lewandowski, Investigating the landscape of physics laboratory instruction across North America, Phys. Rev. Phys. Educ. Res. **16**, 020162 (2020).

[3] A. Karelina and E. Etkina, Acting like a physicist: Student approach study to experimental design, Phys. Rev. ST Phys. Educ. Res. **3**, 020106 (2007).

[4] N. G. Holmes and C. E. Wieman, Introductory physics labs: We can do better, Phys. Today **71**, 38 (2018).

[5] N. G. Holmes, J. Olsen, J. L. Thomas, and C. E. Wieman, Value added or misattributed? A multi-institution study on the educational benefit of labs for reinforcing physics content, Phys. Rev. Phys. Educ. Res. **13**, 010129 (2017).

[6] B. M. Zwickl, D. Hu, N. Finkelstein, and H. Lewandowski, Model-based reasoning in the physics laboratory: Framework and initial results, Phys. Rev. ST Phys. Educ. Res. **11**, 020113 (2015).

[7] B. Pollard, A. Werth, R. Hobbs, and H. Lewandowski, Impact of a course transformation on students' reasoning about measurement uncertainty, Phys. Rev. Phys. Educ. Res. **16**, 020160 (2020).

[8] C. Wieman, Comparative cognitive task analyses of experimental science and instructional laboratory courses, Phys. Teach. **53**, 349 (2015).

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

[10] L. Ivanjek, A. Susac, M. Planinic, A. Andrasevic, and Z. Milin-Sipus, Student reasoning about graphs in different contexts, Phys. Rev. Phys. Educ. Res. **12**, 010106 (2016).

[11] J. D. Gifford and N. D. Finkelstein, Categorical framework for mathematical sense making in physics, Phys. Rev. Phys. Educ. Res. **16**, 020121 (2020).

[12] L. A. Barth-Cohen, S. K. Braden, T. G. Young, and S. Gailey, Reasoning with evidence while modeling: Successes

at the middle school level, Phys. Rev. Phys. Educ. Res. **17**, 020106 (2021).
[13] B. W. Dreyfus, A. Elby, A. Gupta, and E. R. Sohr, Mathematical sense-making in quantum mechanics: An initial peek, Phys. Rev. Phys. Educ. Res. **13**, 020141 (2017).
[14] C. R. Gette, M. Kryjevskaia, M. R. Stetzer, and P. R. Heron, Probing student reasoning approaches through the lens of dual-process theories: A case study in buoyancy, Phys. Rev. Phys. Educ. Res. **14**, 010113 (2018).
[15] J. R. Hoehn and N. D. Finkelstein, Students' flexible use of ontologies and the value of tentative reasoning: Examples of conceptual understanding in three canonical topics of quantum mechanics, Phys. Rev. Phys. Educ. Res. **14**, 010122 (2018).
[16] D. R. Dounas-Frazer, K. L. Van De Bogart, M. R. Stetzer, and H. Lewandowski, Investigating the role of model-based reasoning while troubleshooting an electric circuit, Phys. Rev. Phys. Educ. Res. **12**, 010137 (2016).
[17] E. Etkina, A. Karelina, M. Ruibal-Villasenor, D. Rosengrant, R. Jordan, and C. E. Hmelo-Silver, Design and reflection help students develop scientific abilities: Learning in introductory physics laboratories, J. Learn. Sci. **19**, 54 (2010).
[18] D. R. Dounas-Frazer and H. Lewandowski, The modelling framework for experimental physics: description, development, and applications, Eur. J. Phys. **39**, 064005 (2018).
[19] N. Holmes and C. E. Wieman, Examining and contrasting the cognitive activities engaged in undergraduate research experiences and lab courses, Phys. Rev. Phys. Educ. Res. **12**, 020103 (2016).
[20] T. O. B. Odden and R. S. Russ, Sensemaking epistemic game: A model of student sensemaking processes in introductory physics, Phys. Rev. Phys. Educ. Res. **14**, 020122 (2018).
[21] T. O. B. Odden and R. S. Russ, Defining sensemaking: Bringing clarity to a fragmented theoretical construct, Sci. Educ. **103**, 187 (2019).
[22] A. A. diSessa, Toward an epistemology of physics, Cognit. Instr. **10**, 105 (1993).
[23] D. Hammer, A. Elby, R. E. Scherr, and E. F. Redish, Resources, Framing, and Transfer, in *Transfer of Learning from a Modern Multidisciplinary Perspective* (IAP Publishing, Greenwich, CT, 2005), pp. 187–221.
[24] A. Collins and W. Ferguson, Epistemic forms and epistemic games: Structures and strategies to guide inquiry, Educ. Psychol. **28**, 25 (1993).
[25] 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).
[26] J. Hiebert and P. Lefevre, Conceptual and procedural knowledge in mathematics: An introductory analysis, *Conceptual and Procedural Knowledge: The Case of Mathematics* (Routledge, New York, NY, 1986), pp. 1–27.
[27] J. R. Star, Reconceptualizing procedural knowledge, J. Res. Math. Educ. **36**, 404 (2005).
[28] B. Rittle-Johnson and M. Schneider, Developing conceptual and procedural knowledge of mathematics, *The Oxford Handbook of Numerical Cognition* (Oxford University Press, New York, 2015), pp. 1118–1134.
[29] B. Rittle-Johnson and M. W. Alibali, Conceptual and procedural knowledge of mathematics: Does one lead to the other?, J. Educ. Psychol. **91**, 175 (1999).
[30] B. Modir, J. D. Thompson, and E. C. Sayre, Students' epistemological framing in quantum mechanics problem solving, Phys. Rev. Phys. Educ. Res. **13**, 020108 (2017).
[31] K. S. Serbin, B. J. S. Robayo, J. V. Truman, K. L. Watson, and M. Wawro, Characterizing quantum physics students' conceptual and procedural knowledge of the characteristic equation, J. Math. Behav. **58**, 100777 (2020).
[32] 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).
[33] G. J. Posner, K. A. Strike, P. W. Hewson, and W. A. Gertzog, Accommodation of a scientific conception: Toward a theory of conceptual change, Sci. Educ. **66**, 211 (1982).
[34] D. M. Hammer and E. van Zee, *Seeing the Science in Children's Thinking: Case Studies of Student Inquiry in Physical Science* (Heinemann, London, 2006).
[35] M. J. Ford, A dialogic account of sense-making in scientific argumentation and reasoning, Cognit. Instr. **30**, 207 (2012).
[36] R. S. Russ, R. E. Scherr, D. Hammer, and J. Mikeska, Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science, Sci. Educ. **92**, 499 (2008).
[37] J. Clement, Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics, J. Res. Sci. Teach. **30**, 1241 (1993).
[38] E. Euler, E. Rådahl, and B. Gregorcic, Embodiment in physics learning: A social-semiotic look, Phys. Rev. Phys. Educ. Res. **15**, 010134 (2019).
[39] M. Wilson, Six views of embodied cognition, Psychon. Bull. Rev. **9**, 625 (2002).
[40] R. E. Scherr, H. G. Close, E. W. Close, V. J. Flood, S. B. McKagan, A. D. Robertson, L. Seeley, M. C. Wittmann, and S. Vokos, Negotiating energy dynamics through embodied action in a materially structured environment, Phys. Rev. ST Phys. Educ. Res. **9**, 020105 (2013).
[41] A. Schaafstal, J. M. Schraagen, and M. Van Berl, Cognitive task analysis and innovation of training: The case of structured troubleshooting, Human Factors **42**, 75 (2000).
[42] D. H. Jonassen and W. Hung, Learning to troubleshoot: A new theory-based design architecture, Educ. Psychol. Rev. **18**, 77 (2006).
[43] J. M. May, C. De Grandi, J. M. Gerton, L. Barth-Cohen, A. Beehler, and B. Montoya, Bringing three-dimensional learning to undergraduate physics—insight from an introductory physics laboratory course, Am. J. Phys. **90**, 452 (2022).
[44] American Association for the Advancement of Science, *Vision and Change in Undergraduate Biology Education: A Call to Action* (American Association for the Advancement of Science Press, Washington, DC, 2011).
[45] Howard Hughes Medical Institute—American Association of Medical Colleges Committee, *Scientific Foundations for Future Physicians* (American Association of Medical Colleges, Washington, DC, 2009).

[46] K. Moore, J. Giannini, and W. Losert, Toward better physics labs for future biologists, Am. J. Phys. **82**, 387 (2014).

[47] Office of Budget and Institutional Analysis, Course Enrollment Profile, https://www.obia.utah.edu/data/student-data/course-enrollment-profile/ (2019).

[48] K. L. McNeill and J. S. Krajcik, *Supporting Grade 5-8 Students in Constructing Explanations in Science: The Claim, Evidence, and Reasoning Framework for Talk and Writing* (Pearson, Upper Saddle River, NJ, 2011).

[49] J. P. Walker, V. Sampson, and C. O. Zimmerman, Argument-driven inquiry: An introduction to a new instructional model for use in undergraduate chemistry labs, J. Chem. Educ. **88**, 1048 (2011).

[50] L. S. Vygotsky and M. Cole, *Mind in Society: Development of Higher Psychological Processes* (Harvard University Press, Cambridge, MA, 1978).

[51] J. Saldaña, *The Coding Manual for Qualitative Researchers* (Sage, Thousand Oaks, CA, 2015).

[52] J. M. May, L. A. Barth-Cohen, J. M. Gerton, and C. De Grandi, Students' dynamic engagement with experimental data in a physics laboratory setting, in Proceedings of PER Conf. 2020, virtual conference, 10.1119/perc.2020.pr.May.

[53] M. B. Miles, A. M. Huberman, and J. Saldaña, *Qualitative Data Analysis: A Methods Sourcebook* (Sage Publications, Thousand Oaks, CA, 2018).

[54] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas. **20**, 37 (1960).

[55] J. W. Creswell and C. N. Poth, *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (Sage, Thousand Oaks, CA, 2016).

[56] E. Ochs and B. Schieffelin, Transcription as theory, *Developmental Pragmatics* (Academic Press, New York, 1979), pp. 43–72.

[57] J. Sidnell and T. Stivers, *The Handbook of Conversation Analysis* (John Wiley & Sons, New York, NY, 2012).

[58] G. Jefferson, Glossary of transcript symbols, *Conversation Analysis: Studies from the First Generation* (John Benjamins Publications, New York, 2004), pp. 24–31.

[59] E. Kärkkäinen, The role of *I guess* in conversational stancetaking, in *Stancetaking in Discourse: Subjectivity, evaluation, interaction* (John Benjamins Publishing, New York, 2007), pp. 183–219.

[60] R. E. Stake, *The Art of Case Study Research* (Sage, Thousand Oaks, CA, 1995).

[61] N. K. Denzin, *Interpretive Interactionism* (Sage, Thousand Oaks, CA, 2001).

[62] M. Braaten and M. Windschitl, Working toward a stronger conceptualization of scientific explanation for science education, Sci. Educ. **95**, 639 (2011).

[63] B. Cai, L. A. Mainhood, R. Groome, C. Laverty, and A. McLean, Student behavior in undergraduate physics laboratories: Designing experiments, Phys. Rev. Phys. Educ. Res. **17**, 020109 (2021).

[64] C. E. Wieman and N. G. Holmes, Measuring the impact of introductory physics labs on learning and critical thinking, in *2015 Physics Education Research Conference Proceedings* (American Association of Physics Teachers, College Park, MD, 2015).

[65] K. M. Fisher, K. S. Williams, and J. E. Lineback, Osmosis and diffusion conceptual assessment, CBE Life Sci. Educ. **10**, 418 (2011).

[66] A. Gauthier, S. Jantzen, G. McGill, and J. Jenkinson, Molecular concepts adaptive assessment (MCAA) characterizes undergraduate misconceptions about molecular emergence, CBE Life Sci. Educ. **18**, ar4 (2019).

[67] C. J. Weiss, Scientific computing for chemists: An undergraduate course in simulations, data processing, and visualization, J. Chem. Educ. **94**, 592 (2017).

[68] J. Kottonau, An interactive computer model for improved student understanding of random particle motion and osmosis, J. Chem. Educ. **88**, 772 (2011).