Students' dynamic engagement with experimental data in a physics laboratory setting

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

[1]*Department of Physics & Astronomy, University of Utah, 115 S. 1400 E., Salt Lake City, UT 84112*
[2]*Department of Educational Psychology, University of Utah,*
1721 Campus Center Drive, SAEC 3220, Salt Lake City, UT 84112

Increased emphasis has been given to students' engagement with experimental data as reform efforts have continued to transform the landscape of introductory physics laboratory courses by providing greater opportunities for authentic scientific inquiry and student agency. As a result, students become the primary driving forces of their own experimentation, and the manner in which they engage with experimental data becomes more dynamic and nonlinear. This study presents ongoing efforts to illuminate the nuanced ways students enact various data-based actions when engaging in physics laboratory experiments. In this paper, we present a single case-study analysis of a student group engaging in an inquiry-based physics laboratory to highlight the dynamic and iterative ways the group shifts between multiple data-based actions when expected to be engaging in a single laboratory task. Research data comes from audio and video files of students' computers while they engaged in lab experimentation, coded using a constructivist grounded theory approach to identify multiple data-based actions performed by the students. Results of this case study show that students oftentimes shift between multiple data-based actions on short timescales and that these shifts can take place with implicit iterative patterns, even when the instructional setting is structured for a single experimental task.

I. INTRODUCTION

As the STEM workplace becomes more reliant on data and data analytics, stakeholders invested in undergraduate students have begun advocating for future STEM professionals to obtain greater proficiency with data-based practices (e.g. managing and analyzing data sets, differentiating anomalous data points from normal statistical scatter) during their undergraduate careers [1, 2]. As a result, experimental data has received greater attention in undergraduate STEM curriculum. Within the undergraduate physics community, introductory physics laboratory courses serving physics and non-physics majors alike have increasingly incorporated experiments that prompt students to collect and analyze complex quantitative data to extract meaning about authentic scientific phenomena [3, 4]. Many of these shifts have also resulted in more opportunities for students to engage in authentic scientific inquiry [5, 6]. Opening physics lab courses to inquiry-based instruction can, in turn, heighten student agency, allowing students opportunities to make more decisions during experimentation rather than follow prescribed steps [7–9].

Authentic scientific inquiry is more dynamic and nuanced than the rigid scientific method commonly presented in traditional K-12 and undergraduate instruction [10], especially when students have more agency in their inquiry. This shift towards more authentic scientific inquiry is in line with recent trends at the K-12 level which advocate for a "science-as-practice" approach [11–16]. As students make decisions about how to collect, analyze, and interpret their experimental data, they oftentimes divert from the traditional linear progression of experimental activities. While students in traditional physics labs are commonly expected to engage with experimental data in a linear fashion – from data collection, to analysis, to interpretation, to extracting results, with lab time allocated to each experimental step [17, 18] – these shifts toward more authentic scientific experiences, where students can experimentally act with their own agency, result in students' less procedural, less linear, and more authentic engagement with experimental data.

In order to more effectively develop curriculum for physics courses that provides students authentic experiences with experimental data, what is needed is a more robust understanding of the data-based actions students utilize and how this utilization occurs. In this paper, we consider data-based actions to be the processes and practices, either physical (e.g. manipulating lab equipment) or verbal (e.g. engaging in dialogue to prepare group's collection of data), that students engage with when working with experimental data. This new understanding will influence curricular efforts to further implement authentic experiences with experimental data and because it may shed light on student inquiry more broadly when engaged with authentic data-based actions.

To this end, we ask the research questions: 1) What data-based actions do students engage in when working with experimental data in an introductory physics laboratory course? and 2) In what ways do students shift between data-based actions while working with experimental data in this setting?

To answer these research questions, this paper presents a case study of a student group's dynamic engagement with data-based actions while working with experimental data in a reformed introductory physics laboratory course. The results highlight the dynamic and nonlinear manner of students' utilization of data-based actions during a short period of laboratory activities when students had been ostensibly instructed to engage in a single experimental task. That is, the results capture that even when a single data-based action was expected from an instructional standpoint, students engaged in multiple data-based actions in dynamic and non-linear ways.

II. RESEARCH DESIGN

A. Setting

This study was conducted in a reformed Introductory Physics for Life Sciences (IPLS) laboratory course [4, 5, 19] at a large research intensive university in the western United States. This course was chosen intentionally due to a) its incorporation of 21st century equipment and software which foster complex quantitative experimental data; and b) its pedagogical focus on prompting students to utilize their own agency during experimentation. Students enrolled are primarily life science majors (e.g. pre-medical, biology, etc.) in their final two years of undergraduate studies.

In this course, groups of four (or three) students design their own experiment based on an open-ended guiding prompt with support from teaching and learning assistants. Students have many opportunities to make decisions during their experiments, including what data they collect and in what ways, how they analyze, represent, and interpret that data, how they go about answering their research question, etc. While students maintain considerable agency in their experimentation, the class time is semi-structured such that students have periods of time specifically allocated for different experimental tasks – practice with lab equipment and techniques, experimental design planning, experimentation, group presentations, and writing individual lab reports. For example, in Lab 1 students follow a semi-structured timeline as follows: **Week 1:** Lab Warm-up Activities (45 minutes), Lab Introduction (15 minutes), Experimental Design Planning (20 minutes), Initial Data Collection (20 minutes), Finalize Data Collection and Begin Analysis (40 minutes); **Week 2:** Complete Remaining Experimentation (60 minutes), Group Presentations (45 minutes), Draft Reports and Peer Review (75 minutes). Each lab experiment contains similar structured timing with slight variations. Lab experiment guiding prompts include topics such as biological kinematics, diffusion in living systems, hemodynamics, and modeling of axon signal transport. Students primarily work on lab computers, with complementary equipment such as optical microscopes and spectrometers, to collect and analyze data.

B. Methods

We utilized a constructivist grounded theory approach to identify data-based actions students commonly enact when engaging with experimental data [20, 21]. Video screen capture and associated audio data were collected from lab computers used by fourteen groups across all four of the course's multi-week lab sequences, resulting in roughly one-hundred hours of video screen capture and audio data. Video screen capture and audio data was used for this study because lab computers are the primary means by which students engage with experimentation and data in this course.

To start, a selection of screen capture and audio data was open coded to identify fine-grained instances that contained students' different types of engagement with their group's experimental data. During open coding, particular attention was given to student actions on the computer or with lab equipment, as well as their verbalization of experimental steps. These fine-grained instances were iteratively added upon through coding of additional selections of observational data until broader core categories became salient; the fine-grained instances were then grouped into core categories. These categories were iteratively cross-referenced with additional selections of observational data to reach saturation and minimize potential overlap between categories [21, 22].

The resultant code book consists of ***Data Collection (DCo)***, ***Data Cleaning (DCl)***, ***Data Organization (DO)***, ***Data Manipulation (DM)***, ***Data Representation (DR)***, ***Hypothesizing (Hyp)***, ***Experimental Planning (EP)***, and ***Interpretation (Int)***. These are the broad data-based actions that students are observed to engage in when working with experimental data in this setting; the code book is presented in Table I with definitions and examples. Because of the broad nature of these data-based actions and the population from which this research data comes, this list is not meant to be exhaustive, but is rather meant to provide a useful instrument to support our analysis and answer our research questions.

A feature of this set of data-based actions is its hierarchical structure; students engage in some of the data-based actions in reference to other actions. When students engage in the data-based actions ***Hyp***, ***EP***, and ***Int***, they typically do so while referencing another data-based action that is either forthcoming or had previously occurred. Thus, when ***Hyp***, ***EP***, and ***Int*** are coded in reference to another data-based action, the code is listed as ***Primary Action - XX***, where ***XX*** is the secondary action. For example, when a student engages in ***EP*** while planning steps their group will take to collect experimental data (***DCo***), this is coded as ***EP - DCo***.

Upon completion of this code book, we then conducted a single case-study analysis of a student-group to better understand how students utilize and shift between data-based actions when engaging with experimental data. The case involved is a group of four students (Students A, B, C, and D) engaging in an experiment to analyze zebrafish kinematics; they focused on the social dynamics of zebrafish in a tank by analyzing distances and speeds while the fish swam to then deduce social behaviors (they were provided videos of zebrafish swimming in a small tank). The group was chosen specifically due to: a) their equitable discussions regarding the experimentation (one student did not dominate the conversation or decision making); and b) they were not considered an exemplary group in the course – their lab grades, engagement in class, or manner of experimentation were not noteworthy compared to other groups. Based on the screen capture and audio data and the group's interactions with instructional team members, it appears that the group was on task and doing what was expected of them during class time.

The focus of our analysis is a roughly fifteen minute segment of class time devoted to collecting experimental data. This episode of the group's experimentation was specifically chosen because it occurs at the beginning of the "Initial Data Collection" phase of class time, and from preliminary analysis, this episode appeared particularly dense with many data-based actions apparent. During this portion of class time, students are expected to have finished developing their design plans – an overview of their research question and methods to collect and analyze data – and begin collecting data for their experiment. We coded student utterances and physical actions on the computer using the enumerated data-based actions to develop a deeper understanding of the ways students engage with experimental data, specifically during a segment of class time that aligns with a single data-based action. For our study, an utterance can range from a single word to a series of statements that explicitly encompass the same data-based action. Once a student's comment extends beyond multiple statements, the comment's underlying data-based action shifts, or the speaker changes, a new utterance is defined.

As not all students were working on the lab computer at the same time, inferential coding took place based on student utterances to code multiple students' actions happening simultaneously, or when students' data-based actions were taking place away from the lab computer. For example, in the episode Student D states, "And now two-point-five divided by seventy-one- ..." while Student A is using the lab computer to interpret the group's data collection parameters. We interpreted Student D's statement as their enactment of ***DM*** because it was apparent from the group's dialogue that they were performing a calculation of the group's data using alternate means (e.g. handheld calculator, personal computer).

III. RESULTS

Figure 1 shows the sequence of students' enacted data-based actions during the episode. There are several noteworthy aspects of this student group's enactment of data-based actions. First, though this episode comes from a segment of class time devoted to "Initial Data Collection," the group is primarily engaged with ***EP*** rather than ***DCo***; they were planning for future enactment of data-based actions rather than carrying out those actions themselves. Before this episode, the group completed and submitted their experimental design

TABLE I. Data-Based Actions.

Code	Definition	Example of Code (Action) from Case Study Group
Data Collection (**DCo**)	Enacting steps to collect experimental data using equipment, computer software, etc.	Student A uses computer software with designated settings and parameters to manually track zebrafish swimming by clicking on the zebrafish frame-by-frame.
Data Cleaning (**DCl**)	Identifying and mitigating artifacts (e.g. glitches) in the data that are misaligned with expectations or hypotheses of the experiment of experimental step, often due to equipment or software bugs	Students recognize that the zebrafish video glitches by observing the zebrafish "bounces back" instead of swims forward. **Student A:** "Oh, there's just a glitch in the video ..." **Student B:** "Yeah." **Student A:** "... where it bounces back ..."
Data Organization (**DO**)	Modifying the location, orientation, or other arrangements of experimental data	After receiving position-tracking data from five zebrafish, Student A uses Excel spreadsheet software to compile each zebrafish's x-y coordinate data into columns and place all five data sets next to each other on a single spreadsheet.
Data Manipulation (**DM**)	Performing calculations to transform experimental data from a raw state to a state appropriate for further manipulation, representation, or interpretation	Student A performs calculation to determine the distance-to-pixel ratio for the zebrafish to aid in scaling the position-tracking data from pixels to centimeters. **Student A:** "So twenty-five thousand divided by seventy-one-point-one equal three-hundred and fifty-one-point-six-two."
Data Representation (**DR**)	Creating representations (numerical, graphical, tabular, etc.)	Students create a scatterplot of fish swimming versus distance between fish to interpret relationship between two variables.
Hypothesizing (**Hyp**)	Developing and discussing initial hypotheses of experiment as a whole or future data-based actions (coded in conjunction with other actions, represented as "Hyp - XX")	Student A hypothesizes that there are going to be thirty data points that result from collecting position-tracking data of swimming zebrafish. (**Hyp - DCo**) **Student A:** "Are there gonna be thirty ... there's gonna be, like, thirty points, right?"
Experimental Planning (**EP**)	Discussion of steps to be taken when enacting future data-based actions (coded in conjunction with other actions, represented as "EP - XX")	Students discuss what data they should collect for their experiment. (**EP - DCo**) **Student D:** "I think, okay, I think we should track, that's a good point, I think we should also record the distance from the nearest ..." **Student C:** "'Cause, like, distance is, like, relat, like, I don't know, dist, our distance measurement should be from the closest fish."
Interpretation (**Int**)	Assessing the result of previously enacted data-based actions to test experimental hypotheses, generate experimental claims, extract meaning and results to explain phenomena, etc. (coded in conjunction with other actions, represented as "Int - XX")	Students A and D assess resulting raw data from position-tracking a zebrafish to verify that data collection was carried out properly. (**Int - DCo**) **Student D:** "Let's just open it first, before you like ... like open it in Excel. Fish A ..." **Student A:** \<Opens file in Excel\> **Student A:** "Right? This is right? This looks right?" **Student D:** "Yeah, that is our stuff."

plan prior to moving onto the "Initial Data Collection" phase of class time. However, they spent considerable time during this episode planning for multiple future data-based actions (see Figure 2). While this extensive planning during the "Initial Data Collection" phase could be due to several factors, it may suggest that students iteratively plan throughout experimentation rather than developing a formal plan beforehand, even when time is explicitly given for such formal planning.

Also noteworthy is an emerging pattern of co-occurrences of some data-based actions and not others. For instance, when **DCl** is coded from utterances 229 to 261, the group is identifying and mitigating a glitch in the zebrafish video that they consider has potential to cause significant issues in their data collection and analysis; during this segment of time, the group minimally engages in other data-based actions. This also occurs from utterances 263 to 293 (**DCo**) and utterances 110 to 125 (**DM**), among other instances. Generally, this longer-lasting, stand-alone feature is common for **DCo**, **DCl**, **DO**, and **DM** in these data. This is in contrast to **Hyp**, **EP**, and **Int**, which commonly occur with shorter lengths of time per action, as well as being more significantly intermixed with other actions. Because **DR** was not enacted in this episode, we refrain from stating that it falls into either categorization.

Finally, the student group engaged in **EP** and **Int** in preparation for and in reference to different data-based actions, not solely **DCo** as implied by the class time's intended structure. Because of the data-based actions' hierarchical structure, **EP** and **Int** were coded with secondary actions (see II B). Figure 2 shows the whole-group breakdown of other data-based actions students referred to when enacting **EP**, **Int**, and **Hyp**.
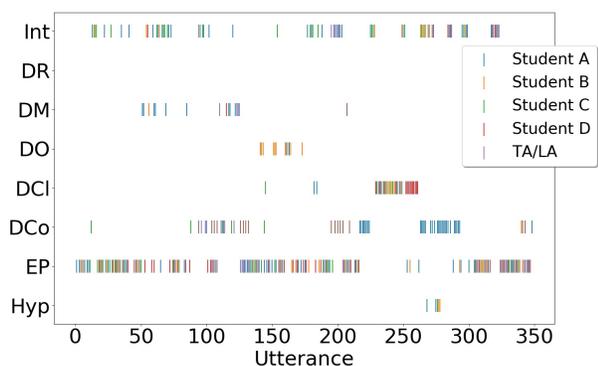
FIG. 1. Full Group Data-Based Actions. Each utterance or action is represented individually on the figure in chronological order, with colors used to differentiate students and teaching/learning assistants.
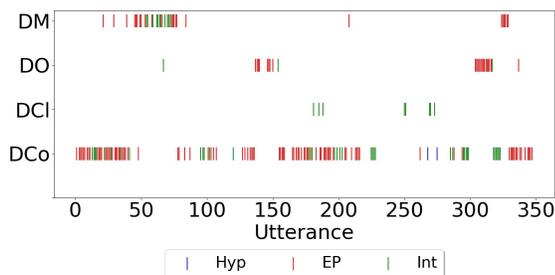

FIG. 2. Secondary Actions. This figure shows the breakdown of data-based actions *Hypothesizing*, *Experimental Planning*, and *Interpretation* in reference to other data-based actions.

By comparing Figures 1 and 2, we begin to see preliminary iterative patterns of *EP* in preparation of a future data-based action, then enactment of the data-based action, then *Int* of the results of the data-based action. For instance, during utterances 1 to 50, the group is primarily enacting *EP-DCo* (see Figure 2) by discussing what parameters they need to adjust on the computer software to collect accurate position-tracking data for the zebrafish. Then, beginning at utterance 88 until utterance 121 they engage in *DCo* (see Figure 1) by measuring and entering into the computer software their parameters with support from the teaching assistant. During this time-frame, the group shifts several times between *DCo* and *Int-DCo*, assessing their measurements and resultant parameters entered into the software. This cycle repeats later when, beginning at utterance 155, the group again enacts *EP-DCo* (see Figure 2) by discussing each group member's responsibilities during future *DCo* steps, what data they need to collect to answer their research question, and what features of the software are necessary to capture this data. Then, from utterances 217 to 293 the group is primarily engaged in *DCo* (see Figure 1), when Student A is using the computer software to man-

ually track a zebrafish to obtain its x-y coordinate data as a function of time. During and after this *DCo*, the group engages in *Int-DCo* (utterances 284 to 323, see Figure 2) by verifying that the resultant data was in the expected format and units, and that the amount of data collected matched the length of the zebrafish video. Overall, emerging from this case is an iterative pattern of student engagement in *EP* in preparation for another data-based action, then enactment of that action, then *Int* of the results of that action; this occurs multiple times during this episode at short timescales.

IV. DISCUSSION AND CONCLUSIONS

This study suggests that students dynamically engage in data-based actions when working with experimental data in introductory physics labs. The case study described above sheds light on the dynamic manner in which students engage in data-based actions, even within a predefined time period in the classroom devoted to a single aspect of experimentation. Notably, the group represented in this case study were observed to iteratively shift between multiple data-based actions in a recurring fashion.

This study is not without its limitations. For example, while the choice to focus on a 15 minute episode during the "Initial Data Collection" phase of class time was intentional, it limits the scope of the results and does not allow for generalization of the dynamic nature of student enactment of data-based actions into other segments of physics laboratories, such as when students complete data analysis and begin formulating their experimental results.

These results also raise new questions that could potentially serve as motivation for future research endeavors. Do these dynamic shifts in student enactment of data-based actions occur throughout an entire physics experiment? What other iterative patterns are prominent during student enactment of data-based actions? What causes students or groups to shift between data-based actions? What timescales are pertinent to these shifts? How do data-based actions align with their summative assessment submissions (i.e. lab reports)? We contend that further exploration of the complexities of students' engagement with experimental data in these settings can elicit further knowledge that can positively influence continued reform efforts in physics classrooms and STEM classrooms more broadly (e.g. [5, 6, 23, 24]).

[1] American Association for the Advancement of Science *et al.*, Vision and change in undergraduate biology education: A call to action, Washington, DC (2011).

[2] AAMC-HMMI Committee *et al.*, Scientific foundations for future physicians. report of the AAMC-HHMI committee (2014).

[3] P. W. Irving, M. J. Obsniuk, and M. D. Caballero, P3: a practice focused learning environment, Eur. J. Phys. **38**, 055701 (2017).

[4] E. F. Redish, C. Bauer, K. L. Carleton, T. J. Cooke, M. Cooper, C. H. Crouch, B. W. Dreyfus, B. D. Geller, J. Giannini, J. S. Gouvea, M. W. Klymkowsky, W. Losert, K. Moore, J. Presson, V. Sawtelle, K. V. Thompson, C. Turpen, and R. K. P. Zia, NEXUS/Physics: An interdisciplinary repurposing of physics for biologists, Am. J. Phys **82**, 368 (2014).

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

[6] Crouch, C. H. and Heller, K., Introductory physics in biological context: An approach to improve introductory physics for life science students, Am. J. Phys **82**, 378 (2014).

[7] D. I. Hanauer, J. Frederick, B. Fotinakes, and S. A. Strobel, Linguistic analysis of project ownership for undergraduate research experiences, CBE - L. Sci. Educ. **11**, 378 (2012).

[8] D. R. Dounas-Frazer, J. T. Stanley, and H. Lewandowski, Student ownership of projects in an upper-division optics laboratory course: A multiple case study of successful experiences, Phys.Rev. Phys. Educ. Res. **13**, 020136 (2017).

[9] N. Holmes, B. Keep, and C. E. Wieman, Developing scientific decision making by structuring and supporting student agency, Physical Review Physics Education Research **16**, 010109 (2020).

[10] X. Tang, J. E. Coffey, A. Elby, and D. M. Levin, The scientific method and scientific inquiry: Tensions in teaching and learning, Sci. Educ **94**, 29 (2010).

[11] National Research Council, *A Framework for K-12 Science Education* (National Academies Press, Washington, DC, 2012).

[12] NGSS Lead States, Next generation science standards: For states, by states, (2013).

[13] Furtak, E. M. and Penuel, W. R., Coming to terms: Addressing the persistence of "hands-on" and other reform terminology in the era of science as practice, Science Education **103**, 167 (2019).

[14] Engle, R. A. and Conant, F. R., Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom, Cognition and instruction **20**, 399 (2002).

[15] K. K. Cetina, *Epistemic cultures: How the sciences make knowledge* (Harvard University Press, 2009).

[16] A. Pickering, *Science as practice and culture* (University of Chicago Press, 1992).

[17] R. Trumper, The physics laboratory–a historical overview and future perspectives, Sci. & Educ. **12**, 645 (2003).

[18] D. Hodson, Re-thinking old ways: Towards a more critical approach to practical work in school science, (1993).

[19] K. B. Marriott, Conference on introductory physics for the life sciences report, (2014).

[20] K. Charmaz, *Constructing grounded theory: A practical guide through qualitative analysis* (Sage, 2006).

[21] Glaser, B. G. and Strauss, A. L., *Discovery of grounded theory: Strategies for qualitative research* (Routledge, 2017).

[22] J. Saldaña, *The coding manual for qualitative researchers* (Sage, 2015).

[23] R. M. Spell, J. A. Guinan, K. R. Miller, and C. W. Beck, Redefining authentic research experiences in introductory biology laboratories and barriers to their implementation, CBE - Life Sciences Education **13**, 102 (2014).

[24] Williams, L. C. and Reddish, M. J., Integrating primary research into the teaching lab: Benefits and impacts of a one-semester cure for physical chemistry, Journal of Chemical Education **95**, 928 (2018).