Developing an Assessment to Investigate Data Analysis in Introductory Physics

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

[1]Department of Educational Psychology, University of Utah, 1721 Campus Center Drive, SAEC 3220, Salt Lake City, UT 84112

[2]Department of Physics & Astronomy, University of Utah, 115 S. 1400 E., Salt Lake City, UT 84112

Abstract:

Recent reforms in science education such as the Next Generation Science Standards have inspired course reforms at both the K-12 and undergraduate level. However, undergraduate science education occurs in a different instructional format than K-12. This distinction, combined with course-specific learning goals, highlights the need for individualized research assessments pertaining to scientific practices. As part of an ongoing project exploring how students enact the practice of Analyzing and Interpreting Data in the Introductory Physics for the Life Sciences (IPLS) labs, we have designed a task-based assessment to investigate the particular data-based actions we have observed students doing in the lab. Our assessment uniquely aims to reflect the open-ended and biologically focused nature of the IPLS lab.

Introduction

The "The Framework for K-12 Science Education" proposes that science learning at the K-12 level has three dimensions: disciplinary core ideas, which comprise key concepts within each discipline; scientific practices, which are tools used to investigate the world (table 1); and cross-cutting concepts, which are reasoning tools that link the scientific disciplines (e.g. patterns, stability and change, energy, etc.)[1]. The subsequent Next Generation Science Standards (NGSS)[2], in conjunction with reports in biology and pre-medicine education calling for authentic research experience[3], have inspired course reforms at both the K-12 and undergraduate levels. Although NGSS originated and is mostly researched in K-12 education, there is a growing interest in applying NGSS to higher education.

However, there are important differences in science education, particularly the lecture/laboratory format of science courses. At the K-12 level, different parts of the class are typically well integrated and taught by a single instructor. In contrast, undergraduate laboratory courses are typically run by graduate teaching assistants, and may not conceptually align with the separately taught lecture. In addition, laboratory courses can vary in structure, ranging from more prescriptive[4] to more inquiry-based where students plan and implement their own experiments[5]. How students engage in scientific practices may look very different depending on the format of the course.

Guidelines influencing science education curriculum and educational research vary from K-12 to undergraduate. Often K-12 education utilizes the NGSS learning progressions to develop curriculum and guide research assessment design. This enables broad use of assessments across different schools[6]. STEM undergraduate curriculum often emphasizes career and further education preparation, yet what learning and research experiences students have at one university may differ from what students experience at another[7], which influences the assessments that researchers design.

For example, the nuances of how physics courses address working with data varies widely, contributing to a variety of assessments. Assessments focusing on data analysis can measure a range of skills such as measurement certainty[8,9,10], interpreting graphs[8,9,10], experimental planning[10], and fitting regression lines to data[9]. Assessment design often aligns to the particular course's established learning goals[8] and may also utilize class observations and interviews with instructors and students to develop the assessment[9]. These methods of assessment design are particularly important for new or reformed courses that have different curriculums from traditional courses.

There has been a growing interest in implementing interdisciplinary physics/biology introductory classes (or physics classes with a biology emphasis for pre-health majors)[11], such as the Introductory Physics for the Life Sciences (IPLS) lecture and laboratory courses birthed out of the NEXUS project at the University of Maryland. These classes aim to teach physics concepts such as fluid dynamics and kinematics within a biological context instead of more traditional physics experiments (e.g. pendulums, inclines, springs)[12]. The University of Utah adopted the two-semester IPLS Laboratories in 2017 with the goal of better serving the pre-medicine student population.

While the NEXUS project did not have the NGSS scientific practices for inspiration in development, the reformed IPLS labs at the University of Utah closely mimic the research process of scientists, roughly aligning with the NGSS Scientific Practices. Instead of providing step-by-step instructions, lab manuals prompt students to ask their own questions about a phenomena, then gather and analyze their own data to answer their questions such as "How has the evolution of the body shape of tuna changed Reynold's number[a predictor of laminar versus turbulent fluid flow]?" Although these labs strongly emphasize the scientific practice of Analyzing and Interpreting Data, there is minimal work on this practice at the undergraduate level compared to the extensive research on other practices like argumentation[13,14,15] or modelling[16,17,18].

Initial research suggests that in this inquiry-based setting, students iteratively engage with experimental data using short timescale data-based actions (table 2)[19]. For example, students must plan an experiment to answer their question of how a fish's body shape effects fluid flow, perhaps utilizing polymer clay and graduated cylinders filled with water (experimental planning). They then video-track different body shape clay models dropped into the graduated cylinder (data collection), and export and process the raw x and y coordinates from the tracking software to calculate Reynold's number (data organization, data manipulation). During experimentation, they may use graphs to visually represent their data to find discrepancies, like catching unchanging positional data from a video glitch (data cleaning). They also use graphs and figures to present their final results to the rest of the class (data representation). Through all of these processes, students are thinking about the results of their actions, such as determining if a data point is a glitch or valid (interpretation), or if the more aerodynamic, ovoid body shape will have a lower Reynold's number and produce laminar flow compared to a more blocky body shape (hypothesizing).

Due to the unique context of the IPLS labs, and the identified data-based actions we want to further explore, no current assessments capture all of the data-based actions we have observed in our labs. In addition, the existing data analysis assessments in physics education require content knowledge[8] and are often multiple-choice[8,9]. Some assessments exist in biology education as well, but bring similar concerns of multiple-choice format[20,21] and needed content knowledge[20]. Since data analysis is understudied in comparison to other practices, there is a need for open-ended assessments to yield more detailed information about student's thought processes, which could then support future development of multiple-choice assessments. Thus, we designed an open-ended, task-based assessment to target the data-based actions students enact in the lab, and do so in a biological context without requiring extensive content knowledge.

Design and Refinement of Data-Based Actions Assessment

To support the development of our assessment, we conducted an initial qualitative analysis of recorded interviews and classroom observations. In the interview, pairs of students worked with a video of a myosin motor moving a microsphere along a microtubule and were asked to develop and enact a plan to characterize the movement of the motor protein. Class observations involved video and audio recordings of students working in the first semester course. Through this we identified and defined the kinds of data-related activities students do in the labs[19], resulting in a set of eight data-based actions (Table 2), and initial ideas of questions that could mimic the interactions with data that students had in the lab.

We then designed assessment items with two requirements. First, the item prompt must be framed in biology to be disciplinarily connected to the students, as the majority of the IPLS student population are pre-health students. Second, the items must clearly focus on particular data-based actions. Due to the paper assessment format, the action "data organization" was not targeted as this action occurs most frequently when exporting/acquiring data from software. Item topics included a variety of biological disciplines (ecology, microbiology, botany, oncology), and cumulatively focused on all of the data-based actions through 24 free-response questions, three of which require a drawing response. These items were organized into eight multi-part items (example item, figure 1; descriptions, table 3). The eight items were split into pre-course and post-course assessment versions based on equitable distribution of data-based action targets, with 25% of the items occurring on both the pre- and post-course assessments (table 4).

To ascertain how the item prompts were interpreted, we conducted think-aloud interviews with faculty experienced in research-based data analytics. Items were refined to address implied biological knowledge assumptions, as this assessment was not intended to assess biology knowledge. We also conducted ten think-aloud interviews with STEM undergraduates and graduates to refine item prompts for clarity. Following the refinement after the think-aloud interviews, an initial pilot of the paper assessment was conducted in the first semester ILPS lab course, with a total of 246 pre-course and post-course responses collected.

Analysis of the responses began with an initial 30% of collected pre-course and post-course assessments. Preliminary analysis focused on the most common response for each item as well as variation in responses. Using these responses and the think-aloud responses, the initial rubric was refined to five levels to better capture the variation. Each level has unique descriptions for each question to address both correct responses and the quality of the explanation given to support the response. As such, level one addresses incorrect responses with poor or no or no support, while level five addresses correct responses with thorough support for the response. While a global rubric would be useful for comparing student responses between prompts[22], the items vary in what students are asked to do, which necessitates an item-specific rubric.

Student responses produced not only variety in response, but interesting contradicting responses. For example, on the item "Plant Growth in Low Light," some students sketch an exponential line for sunflower growth rate (which matches the given data), yet when later asked to pick an equation that best fits the data, they chose a linear equation without explaining why. Think-aloud interviews will be developed further explore these interesting responses, providing students opportunities to describe their thought processes more extensively. This will support the larger goal of developing a deeper theoretical understanding of how students engage in data analysis in an IPLS lab setting.

Discussion

It is important to consider how an assessment was created when considering it as a research tool. Designing a valid assessment involves defining the constructs one wants to assess[23]. Similarly, it is just as important to define what skills or activities one wants to study before picking an assessment, as this determines what assessments will validly work. If the definitions for the skills do not line up, the validity of using that assessment for something it was not designed is questionable. Determining the preferred format limits the applicable assessments.

Upon defining our data-based actions, determining that knowing the student's explanation was important for our research, and that we wanted to keep in line with the IPLS laboratory using biology as a setting, there were limited existing surveys that targeted what we wanted without introducing other concerns. While we could have used one of the physics data analysis assessments[8,9,10] and attempt to overlay our data-based actions, this brings concerns of construct validity as well as using an assessment with a different student population than which the assessment was initially designed[24]. Furthermore, these assessments are framed in a traditional physics context, which would not fit the interdisciplinary context of the IPLS laboratories. The limitations with the current assessments necessitated designing our own.

While designing a unique assessment solves the above issues, creating a quality assessment is not an easy task. Designing and rigorously validating an assessment can be a multi-year process. Consider how much validation is appropriate for your assessment; will it be used in other student populations, at other universities? What is the ultimate purpose of the assessment? Answering these questions will guide the amount and types of validity research needed.

Since the practice of Analyzing and Interpreting Data is understudied at the undergraduate level, our assessment is primarily exploratory. That is, a data collection tool to collect student responses to specific questions, which will then support the development of new interview protocols. As we want to know the thought process behind a student's response, the think-aloud interview format better suits our research interests. Our assessment was not designed to assess "skill level" in these data-based actions, but rather capture how students respond to hypothetical versions of these data-based actions that they do in the lab.

Assessments can be useful tools in a variety of contexts. They can provide an initial expedition into a construct, and they can become refined tools to assist instructors and researchers. It is important to determine the state of current research in a subject, such as data analysis in physics education, before developing and implementing an assessment. While some scientific practices like argumentation[15] and modelling[16] have existing frameworks to describe student thinking, other practices do not. Thus, it is important for researchers to explore the nuances of other scientific practices to build the literature that supports curriculum change.

Funding

This material is based upon work supported by the National Science Foundation Grant No. 1938721.

References

1. National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.
2. NGSS Lead States (2013). *Next generation science standards : by states, for states. 1, The standards - arranged by disciplinary core ideas and by topics*. National Acad. Press.
3. Brewer, C. A., & Smith, D. (2011). Vision and change in undergraduate biology education: a call to action. *American Association for the Advancement of Science, Washington, DC*.
4. Trumper, R. (2003). The physics laboratory–a historical overview and future perspectives. *Science & Education*, *12*(7), 645-670.
5. Moore, K., Giannini, J., & Losert, W. (2014). Toward better physics labs for future biologists. *American Journal of Physics*, *82*(5), 387-393.
6. Knight, A. M., Alves, C. B., Cannady, M. A., McNeill, K. L., Pearson, P. D., & Pearson, P. D. (2014). Assessing middle school students' abilities to critique scientific evidence. In *annual meeting of NARST, Pittsburg, PA*.Brewer, C. A., & Smith, D. (2011). Vision and change in undergraduate biology education: a call to action. *American Association for the Advancement of Science, Washington, DC*.
7. Brew, A. (2013). Understanding the scope of undergraduate research: A framework for curricular and pedagogical decision-making. *Higher Education*, *66*(5), 603-618.
8. Eshach, H., & Kukliansky, I. (2016). Developing of an instrument for assessing students' data analysis skills in the undergraduate physics laboratory. *Canadian Journal of Physics*, *94*(11), 1205-1215.
9. Day, J., & Bonn, D. (2011). Development of the concise data processing assessment. *Physical Review Special Topics-Physics Education Research*, *7*(1), 010114.
10. Kung, R. L., & Linder, C. (2006). University students' ideas about data processing and data comparison in a physics laboratory course. *Nordic studies in science education*, *2*(2), 40-53
11. Moore, K., Giannini, J., & Losert, W. (2014). Toward better physics labs for future biologists. *American Journal of Physics*, *82*(5), 387-393.
12. Redish, E. F., Bauer, C., Carleton, K. L., Cooke, T. J., Cooper, M., Crouch, C. H., ... & Klymkowsky, M. W. (2014). NEXUS/Physics: An interdisciplinary repurposing of physics for biologists. *American Journal of Physics*, *82*(5), 368-377.

13. Figueira, M. J. S., Nardi, R., & Cortela, B. S. C. (2019, August). Introducing scientific argumentation practices in physics teacher's undergraduate curricula. In *Journal of Physics: Conference Series* (Vol. 1286, No. 1, p. 012038). IOP Publishing.
14. Lammers, A., Goedhart, M. J., & Avraamidou, L. (2019). Reading and synthesising science texts using a scientific argumentation model by undergraduate biology students. *International Journal of Science Education*, *41*(16), 2323-2346.
15. Rapanta, C., & Christodoulou, A. (2019). Walton's types of argumentation dialogues as classroom discourse sequences. *Learning, Culture and Social Interaction*, 100352.
16. Zwickl, B. M., Hu, D., Finkelstein, N., & Lewandowski, H. J. (2015). Model-based reasoning in the physics laboratory: Framework and initial results. *Physical Review Special Topics-Physics Education Research*, *11*(2), 020113.
17. Dounas-Frazer, D. R., & Lewandowski, H. J. (2018). The Modelling Framework for Experimental Physics: description, development, and applications. *European Journal of Physics*, *39*(6), 064005.
18. Bierema, A. M. K., Schwarz, C. V., & Stoltzfus, J. R. (2017). Engaging undergraduate biology students in scientific modeling: Analysis of group interactions, sense-making, and justification. *CBE—Life Sciences Education*, *16*(4), ar68.
19. May, J. M., Barth-Cohen, L. A., Gerton, J. M., De Grandi, C. (2020). Student's dynamic engagement with experimental data in a physics laboratory setting. *PERC Proceedings 2020.*
20. Rybarczyk, B. J., Walton, K. L., & Grillo, W. H. (2014). The development and implementation of an instrument to assess students' data analysis skills in molecular biology. Journal of microbiology & biology education, 15(2), 259.
21. Hurney, C. A., Brown, J., Griscom, H. P., Kancler, E., Wigtil, C. J., & Sundre, D. (2011). Closing the loop: Involving faculty in the assessment of scientific and quantitative reasoning skills of biology majors. *Journal of College Science Teaching*, *40*(6), 18-23.
22. Hull, M. M., Kuo, E., Gupta, A., & Elby, A. (2013). Problem-solving rubrics revisited: Attending to the blending of informal conceptual and formal mathematical reasoning. *Physical Review Special Topics-Physics Education Research*, *9*(1), 010105.
23. Wilson, M. (2005). *Constructing measures: An item response modeling approach.* Mahwah, NJ: Lawrence Erlbaum Associates.
24. Hsu, C. C. & Sanford, B. A. (2010) Instrumentation. In Salkind, N. J., ed. *Encyclopedia of Research Design* (1st edition, Volume 2, pgs. 607-610). Thousand Oaks(CA): SAGE Publications, Inc.

Figures and Tables

Table 1: NGSS Scientific Practices

Scientific Practice	Definition
Asking Questions	"…Formulating empirically answerable questions about phenomena, establishing what is already known, and determining what questions have yet to be satisfactorily answered."
Developing and Using Models	"…The construction and use of a wide variety of models and simulations to help develop explanations about natural phenomena."
Planning and Carrying out Investigations	"…Planning and carrying out a systematic investigation, which requires the identification of what is to be recorded and, if applicable, what are to be treated as the dependent and independent variables (control of variables)."
Analyzing and Interpreting Data	"…use a range of tools—including tabulation, graphical interpretation, visualization, and statistical analysis—to identify the significant features and patterns in the data. Sources of error are identified and the degree of certainty calculated."
Using Mathematics and Computational Thinking	"…[utilizing mathematics and computation] for a range of tasks, such as constructing simulations, statistically analyzing data, and recognizing, expressing, and applying quantitative relationships."
Constructing Explanations	"…explicit applications of theory to a specific situation or phenomenon, perhaps with the intermediary of a theory-based model for the system under study."
Engaging in Argumentation from Evidence	"…identifying the strengths and weaknesses of a line of reasoning and for finding the best explanation for a natural phenomenon."
Obtaining, Evaluating, and Communicating Information	"…the communication of ideas and the results of inquiry—orally, in writing, with the use of tables, diagrams, graphs, and equations, and by engaging in extended discussions with scientific peers."

Originally published in *A Framework for K-12 Science Education*[1]

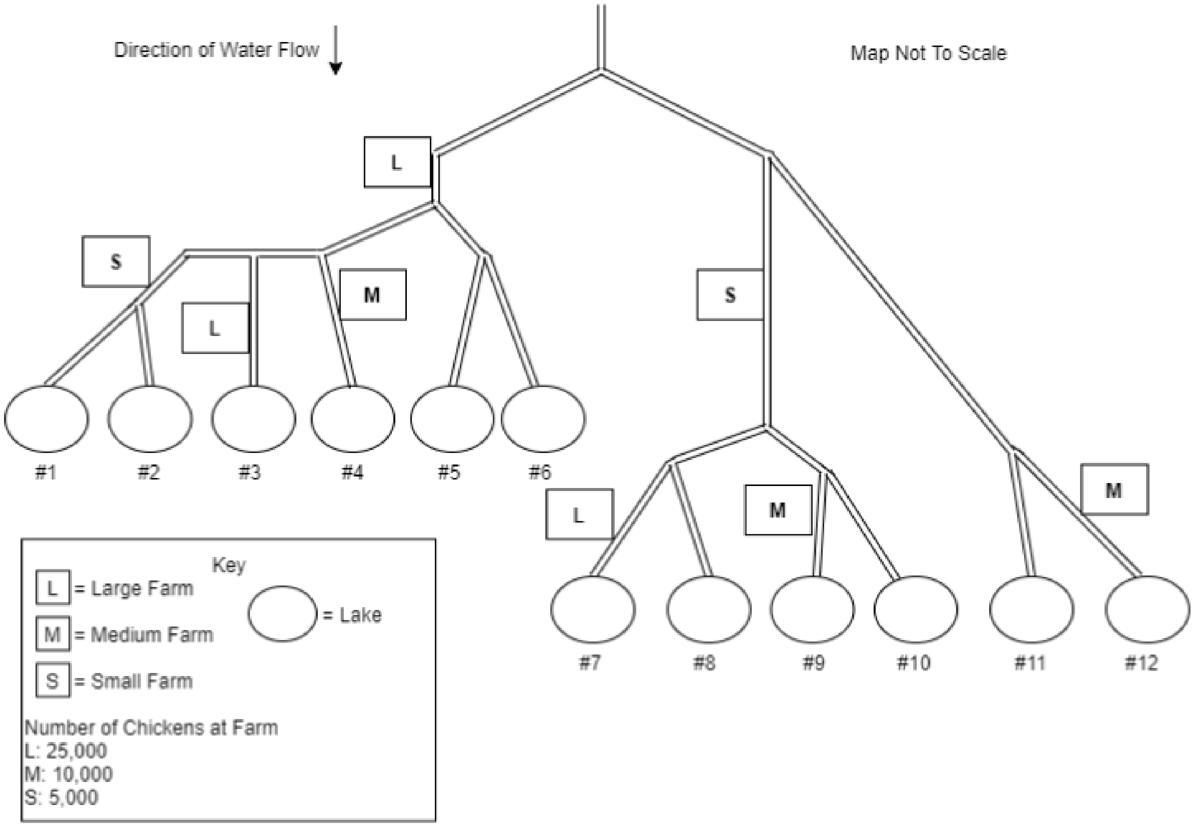
Table 2: Data-Based Actions in IPLS Laboratory Setting

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
Data Collection (*DCo*)	Enacting steps to collect experimental data using equipment, computer software, etc.
Data Manipulation (*DM*)	Performing calculations to transform experimental data from a raw state to a state appropriate for further manipulation, representation, or interpretation
Data Organization (DO)	Modifying the location, orientation, or other arrangements of experimental data
Data Representation (*DR*)	Creating representations (numerical, graphical, tabular, etc.)
Experimental Planning (*EP*)	Discussion of steps to be taken when enacting future data-based actions
Hypothesizing (*Hyp*)	Developing and discussing initial hypotheses of experiment as a whole or future data-based actions
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.

Reproduced from May et al., 2020[16].

Table 3: Descriptions of Task-based Assessment Items

Plant Growth in Low Light	This prompt describes an experiment where sunflowers are only exposed to two hours of sunlight a day. Students must analyze a table of plant heights over 30 days to point out and describe faulty data, sketch the growth trend, compare the data table with a graph of new data of sunflowers grown in full sun, and pick an equation of best fit to the data.
Experimental Critique	A vignette of a student conducting an experiment on California blackworms (small, transparent worms). The student prepares solutions of two substances (nicotine and alcohol) and then tests their effects on the worm's heart rate. There are numerous errors and poor experimental technique, which students are asked to point out and discuss, and provide solutions to these errors.
Microscopic Algae Measurement	A vignette of two students encountering an issue when trying to digitally measure a *Chlamydomonas* alga under the microscope. Students must point out what the cause for the size discrepancy is (compared to other groups' results) and how to resolve the discrepancy.
Cancer Research in Mice	Given a brief description of a hypothetical cancer treatment, and four limitations related to the experiment, students are asked to determine the success of moving trials from mice to humans in light of the limitations, and either argue for the success or discuss what further research must be done prior to human trials.
Predicting Tree Distribution	Given four heat maps depicting soil moisture, soil surface temperature, and soil fertility based on nitrogen concentrations, and the preferences for four tree species, students must place icons where they predict the species will be. Students must modify these placements in light of new information (pH heat map and tree preferences), and discuss if this model can be applied to other scenarios.
Organic Pollution Research	Given a diagram of a waterway system that depicts the locations of chicken farms and lakes, students must determine which lakes to sample to research organic pollution levels caused by contaminated runoff from chicken farms. Then, students choose the microscopic sampling technique they would use to count indicator species of algae, and when given Palmer pollution index values for the lakes, what farms they would close to efficiently deal with the pollution.
Biological Spectroscopy	A vignette of two students attempting to use a fiber-optic based spectrometer to sample the fluorescence of a bacteria sample expressing green fluorescent protein. Students must pick what filters would best mitigate extraneous light sources (given their spectra), and other solutions to ensure they only observe their sample's fluorescence.
Species Population Relationships	Given a brief description of predator/prey and temporal ecological relationships, and a description of a lake where herons, crawfish, and diatom populations have been measured over several years, students must describe what relationships they see in the graph, and apply that information to describe the future population trends in a scenario where the heron population has a sudden large increase.



a. Based on the information provided in the map, which lakes (no more than 6) would you select to representatively sample all of the lakes in this waterway system? Select the lakes on the map. Why did you choose those lakes?

Figure 1: Organic Pollution Research, paper version
After an initial prompt describing organic pollution entering a waterway system via rainwater runoff contaminated with chicken feces from chicken farms, students are told their research team can only survey six lakes and must pick which lakes they must survey.

Table 4: Target Data-Based Actions of each Assessment Item

Pre-Course Assessment

Item	DCl	DCo	DM	DR	EP	Hyp	Int
Plant Growth in Low Light	X		X	X			X
Microscopic Algae Measurement		X	X		X		X
Experimental Critique					X		X
Cancer Research in Mice					X	X	X
Predicting Tree Distribution		X	X	X	X	X	X

Post-Course Assessment

Item	DCl	DCo	DM	DR	EP	Hyp	Int
Organic Pollution Research		X			X		X
Plant Growth in Low Light	X		X	X			X
Biological Spectroscopy		X			X		X
Experimental Critique					X		X
Species Population Relationships			X	X		X	X

An “X” indicates a target for an item. Abbreviations: DCl-Data Cleaning; DCo-Data Collection; DM-Data Manipulation; DR- Data Representations; EP-Experimental Planning; Hyp-Hypothesizing; Int-Interpretation