UTOPIA: An Observation Protocol for Measuring Scientific Inquiry in Undergraduate Laboratory Courses

Leanne Lupone
Chemistry

Author

Leanne Lupone became involved in a research group called Project SPROUT, which examined instructional practices in large undergraduate lecture courses at UC Irvine, particularly in STEM (science, technology, engineering, and mathematics) subjects. Based on this work, Leanne decided to focus on investigating best practices in undergraduate lab courses. Her favorite part of the research experience was the potential for this project to redefine STEM education for the next generation of graduates. After graduation, Leanne will be a Fulbright scholar in Uruguay and teach high school science before pursuing a doctorate in curricular studies or educational policy.

Abstract

There has been a recent movement toward inquiry in K-12 classrooms to increase critical, scientific thinking among young learners. However, undergraduate laboratory courses are not held under national standards like K-12 public schools. In order to compare student outcomes from inquiry-based labs with traditional lab courses, this project seeks a way to measure the amount of inquiry in a lab. Such an assessment would help create a standard across the country and evaluate the academic rigor of a science lab curriculum. This project developed a form called the Undergraduate Teaching Observation Protocol for Inquiry Assessment (UTOPIA), based on components of inquiry-based labs: collaboration, discovery, iteration, and use of the scientific method. It was found that inquiry depends strongly on the actions of both students and the lab instructor, making it difficult to differentiate between measuring the observed class and its curriculum. Tests revealed differences depending on whether the observing researcher focused on the intended curriculum or student and instructor interactions. Future research should determine whether the protocol should measure the curriculum or observations, modify UTOPIA to clarify this emphasis, pilot the protocol at multiple universities, and develop methods to transfer qualitative data from UTOPIA into quantifiable inquiry scores for comparison with student outcomes.

Faculty Mentor

In this outstanding study, Leanne Lupone developed the first observational protocol on the use of inquiry-based instruction in college lab courses and then piloted it to examine how it could be further improved. Her highly innovative project made an important contribution to our efforts to develop evidence-based approaches for assessing undergraduate instruction. This is a terrific example of how faculty-mentored undergraduate researchers can go beyond rote tasks to pursue their original ideas in research. I encourage other undergraduates to not only join research projects but also to take a proactive role in them. Your innovation and ideas are needed to help address the social, scientific, and educational challenges faced by our university and nation.

Mark Warschauer
School of Education

Key Terms

- Education
- Evaluation Protocol
- Inquiry
- NGSS Standards
- Traditional Instruction
- Undergraduate Laboratory Courses

Background

Science education across the United States has been changing over recent years with the development and implementation of the Next Generation Science Standards (NGSS). Current teachers are being re-trained to adapt their curriculum to the three dimensions of NGSS (Figure 1). Since this project seeks to use a standardized and widely accepted definition for inquiry, it turns to NGSS and the government's source of research for guiding policies in education, the National Research Council (NRC) to define the components of high-level inquiry in education. The NRC describes inquiry as "asking questions, planning and conducting investigations, using appropriate tools and techniques to gather data, thinking critically and logically about relationships between evidence and explanations, constructing and analyzing alternative explanations, and communicating scientific arguments" (NRC, 2000). While this definition was written 14 years before the development of NGSS, it is almost identical to the Science and Engineering Practices (SEPs) in Figure 1.

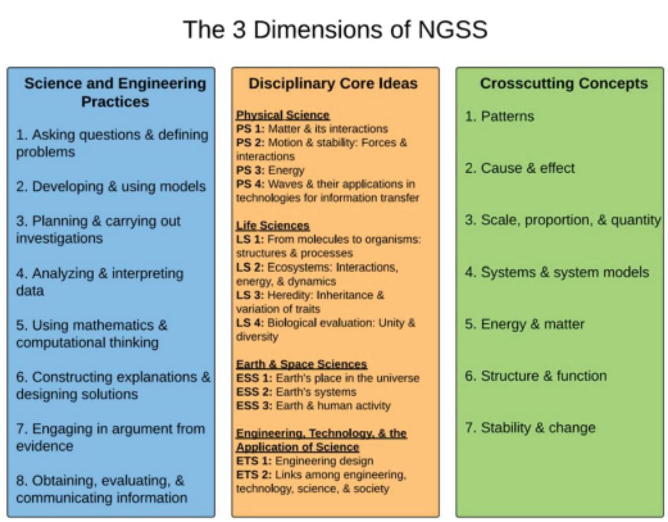

Figure 1

The three dimensions that make up Next Generation Science Standards (https://www.kmajda.net/ngss.html)

While the other two dimensions primarily instruct *what* to teach, the SEPs guide teachers on *how* to teach so that students will exercise thinking as a true scientist. Combined with the other dimensions, the SEPs ensure that inquiry is being used in a classroom (Midwest Comprehensive Center, 2016). The NRC claims that "students at all grade levels and in every domain of science should have the opportunity to use scientific inquiry and develop the ability to think and act in ways associated with inquiry," because science involves revision, such as exploring unexplained phenomena and modifying theories from evidence, not the knowledge we already hold (NSTA, 2018). While NGSS was intended for K-12 classrooms, laboratory courses at universities are indisputably a place where future STEM (Science, Technology, Engineering and Mathematics) workers and researchers should develop inquiry skills; however, these courses have failed to incorporate the SEPs. According to research by Dr. Patrick Brown in undergraduate labs, "college science faculty members held a 'full and open inquiry' view, seeing classroom inquiry as time consuming, unstructured, and student directed." Instead, many courses had students follow a series of specific procedures to collect data and use the data to answer narrow questions with known, correct answers, requiring little critical thinking into the concepts explored by the experiment. These types of experiments are called "cookbook" labs because the procedure is like a recipe, followed without deviation toward one intended, correct product.

Literature Review

This project's goal is to develop a protocol to measure the level of inquiry in an undergraduate laboratory course. Research reveals the importance of this endeavor by comparing results from traditional and inquiry lab instruction. To determine how to measure inquiry, it must also be understood what makes a lab "inquiry-based." This literature review introduces the necessary elements of inquiry and previously designed evaluation tools adapted for this project.

Previous Assessments of Undergraduate Labs

Defining Inquiry-Based Labs and their Components: In 1996, the NRC published the *National Science Education Standards*, in which they defined "inquiry" as "the diverse ways in which scientists study the natural world and propose explanations based on the evidence derived from their work" (NRC, 1996). An analysis of these standards by the Midwest Comprehensive Center (MCC) emphasizes the significance of this definition: "Rather than inquiry being defined as an exclusively hands-on process, or set of rigid and prescribed steps to be followed, the NRC had redefined inquiry as an approach that encompasses both knowledge and skills" (MCC, 2016). This approach is informally known as "hands-on, minds-on." Still, educators continued to have varying interpretations of inquiry instruction, usually assuming that inquiry means learning science through a series of steps, interchangeable with the scientific method. To clarify the misconception, the NRC published *Inquiry and the National Science Education Standards* several years later, specifying "Essential Features of Classroom Inquiry":

1. Learners are engaged by scientifically oriented questions.

2. Learners give priority to evidence, which allows them to develop and evaluate explanations that address scientifically oriented questions.

3. Learners formulate explanations from evidence to address scientifically oriented questions.

4. Learners evaluate their explanations considering alternate explanations, particularly those reflecting scientific understanding.

5. Learners communicate and justify their proposed explanations.

Ten years after this clarifying document was released, researchers Asay and Orgill analyzed more than 300 articles in a science educator journal to learn how authors interpreted inquiry, as defined by the NRC's document. They found that "students collecting data" was mentioned in 82% of articles on inquiry, whereas "formulating explanations from evidence," "connecting explanations to scientific knowledge," and "communicating and justifying explanations" were in fewer than a quarter of the studied articles (Asay & Orgill, 2010). Furthermore, most activities mentioned in the articles were teacher-led instead of student-led. The researchers concluded that "teachers seem to view inquiry more as a process than as a vehicle for learning science content" (Asay & Orgill, 2010). A standardized list of what constitutes inquiry, has not been established due to its many varying degrees.

Several analyses have already been done to categorize inquiry in discrete levels. Research led by Fay evaluated inquiry in undergraduate chemistry labs, similar to this project, using a rubric summarized by Table 1. The group used their rubric to compare the difference in inquiry levels of organic and general lab courses. Their results found that 42% of general chemistry labs were a level 2 or higher, but only 18% of organic chemistry labs were level 2 and none were level 3 (Fay, 2007). Based on this data, the level of inquiry decreased as students moved to more advanced laboratory courses.

Table 1

Rubric for assessing inquiry in undergraduate chemistry labs (Fay 216).

Lvl	Problem/Question	Procedure/Method	Solution
0	Provided to student	Provided to student	Provided to student
1	Provided to student	Provided to student	Constructed by student
2	Provided to student	Constructed by student	Constructed by student
3	Constructed by student	Constructed by student	Constructed by student

A similar method was used by Bruck, as shown in Table 2, to analyze undergraduate labs from several disciplines. Both of these projects simplify the definition of inquiry as "varying degrees of independence" (Bruck, 2008), as shown in their rubrics. Fay's research used both observation of experiments and corresponding lab manuals to categorize the labs, while Bruck's research only analyzed lab manuals from each course to assess inquiry. Bruck's study found that 92% of the 386 analyzed labs were below level 1 and none were higher than level 2. Although both studies strive to categorize labs, their oversimplification of inquiry renders their results unsatisfactory. The NRC (2000) and NGSS (2014) both agree that inquiry is not synonymous with the scientific method, and their method of ranking inquiry based on student independence would be insufficient if used to standardize inquiry across the nation.

Table 2

Rubric for assessing inquiry in undergraduate science labs (Bruck, 2008).

Characteristic	**Level 0: Confirmation**	**Level ½: Structured Inquiry**	**Level 1: Guided Inquiry**	**Level 2: Open Inquiry**	**Level 3: Authentic Inquiry**
Problem/Question	Provided	Provided	Provided	Provided	Not Provided
Theory/Background	Provided	Provided	Provided	Provided	Not Provided
Procedures/Design	Provided	Provided	Provided	Not Provided	Not Provided
Results analysis	Provided	Provided	Not Provided	Not Provided	Not Provided
Results communication	Provided	Not Provided	Not Provided	Not Provided	Not Provided
Conclusions	Provided	Not Provided	Not Provided	Not Provided	Not Provided
	(More Structure)				**(Less Structure)**

Inquiry as Distinguished from Traditional Labs

There is not a particular set of labs that can be classified as inquiry-based. Depending on how the teacher structures a course, almost any experiment can be converted into an inquiry-based lab. For example, a lab on DNA extraction can be performed with a given set of instructions. The students follow the procedure and complete the lab by extracting DNA. However, they did not use collaboration, relate content to the lab, or understand the purpose of each step in the procedure. This would be an example of the cookbook method. According to Alozie, "Scientific inquiry involves engaging learners in scientific practices such as asking scientific questions, experiencing phenomena by designing and conducting investigations, collecting and analyzing data, constructing explanations based on evidence, and sharing findings with others" (Alozie, 2012). Using this principle, teachers can convert this cookbook lab into an inquiry lab by adding several components. First, they engage students by asking about their previous knowledge of DNA and cell structure and elicit questions they might want to answer. This step fosters curiosity about the content, which would be missing in a set of procedural instructions with no introduction to the scientific background. Next, the teacher gives instructions for how to extract DNA out of order. In order to continue with the exploration step of the lab, students must first think critically about the concepts behind extraction to figure out the correct procedure. This reinforces the content beyond any cookbook lab, since they must scientifically defend their choice without guidance from the teacher. Finally, the students have a discussion. This provides an open-endedness for the students to think individually, get feedback from others, and give insight to the ideas of their peers. If someone points out a mistake, students must go back and fix their experimental design. This critique and revision process closely models that of a real-world research lab. With this inquiry-based method, the students have a better understanding of the content because they work with their minds, rather than having the answers handed to them. The student interactions and critical thinking involved in each step helps students develop problem-solving skills they will need in their future academic and professional endeavors.

Effectiveness of Inquiry-Based Labs and Traditional Instruction

The cookbook method is an informal name for labs described as "hands-on and minds-off; in essence, students are not required to think critically because the lab activities have assumed the cognitive load for the students" (Alozie, 2012). In the past several decades there has been a movement to increase the number of inquiry-based labs, in which "students take greater responsibility for their own learning by designing experimental procedures and communicating their understanding and reasoning" (Alozie, 2012). In other words, inquiry requires students to think critically about the purpose of actions and results in experiments as they relate to scientific concepts.

Another article criticizes the cookbook method's narrow objective of working toward one right answer. Holding, a biology teacher, says that introducing research in this way creates the misconception that science has a "beginning" and "end" point, and that the path between is a straightforward and concrete process. Instead, he prefers an inquiry-based lab to teach the scientific method while concurrently meeting content objectives. This is meant to introduce the process of experimental design and demonstrate the importance of collaboration. In the observed biology class, students discuss what "real science" is and then design an experiment to address a scientific question. Afterwards, students present their findings to the class in a formal poster setting and then compose a literature review on their topic for homework. Holding presents the benefits of using this method: "Finally, the described method of showing students the real, nonlinear way that scientists approach problems could be a valuable experience for future nonscientists by giving a relatable, real-world perspective of a human endeavor that is often clouded by misconceptions" (2014). In this lab, students investigate science beyond answering their research question. They learn how to become scientists through use of the scientific method and collaboration, which helps them develop skills beyond the content of the class.

Previous Tools for Assessing Inquiry

This project builds on the Measure of Effective Teaching (MET) project, which designed an observation protocol to be used in teacher evaluations. The goal of rating teaching practices was to give feedback to teachers and improve student learning. Protocols are necessary for evaluations because scoring must be rigorous, standardized and objective to be effective.

The MET project gives insights into designing or modifying protocols to fit a specific district's standards. If appropriate, a protocol can also be content-specific, as in the case of UTOPIA. In this project, researchers tested for teacher effectiveness, so the research group had to give a comprehensive and widely accepted definition of good teaching practices. In order to ensure scoring is standardized, the MET project suggests observer training and certification for the specific protocol being used. The training gives observers a standardized view of what effective teaching practices look like for accurate and reproducible results. Minimizing bias to allow for objective evaluations was also addressed during observer training (Joe, 2013).

The protocol for the MET project was designed by looking at previously constructed and validated protocols. The team eliminated questions that required making inferences about the teacher, previous knowledge about the course, or using tools other than observation alone, such as reading a lesson plan. The protocol measured components of effective teaching, such as student engagement, questioning techniques, classroom management, and presentation of content.

Considerations in designing a protocol include content-based validity, convergent validity, and generalizability. "Content-based validity" ensures that the protocol is measuring what it is designed to measure. In the MET project, the study focused on evaluating teachers and not the curriculum. "Convergent validity" requires that the results of the protocol correlate as expected with other measured variables, such as a teacher's perception of their performance. Finally, "generalizability" encourages interpretations to be made only if the sampling of observed lessons can be generalized as representative of the evaluated teacher.

Another research study designed a Laboratory Course Assessment Survey (LCAS) to evaluate design features of biology labs. LCAS results were used to differentiate Course-based Undergraduate Research Experiences (CUREs), from traditional lab courses. The purpose of the survey was to "characterize CUREs and link particular CURE design features to specific student outcomes" (Corwin, 2015). The LCAS dimensions included collaboration, discovery, and iteration. Collaboration measures the frequency of engaging with peers and using metacognition. Discovery measures whether the students learn phenomena unknown to themselves and the scientific community. Finally, iteration assesses how much the students are able to revise their theories based on repeating steps for new evidence. These three dimensions were assessed in surveys distributed electronically to students who had taken one of the 16 lab courses participating in the study.

The study identified three concepts necessary for developing a survey-based protocol: dimensionality, reliability, and validity. A survey has dimensionality when several items have correlated responses and are testing for the same constructs. Reliability is the consistency of results across implementations of the survey in similar populations and settings. Finally, a survey has validity when it measures what it is intended to measure. The study tested for three types of validity: construct validity, face validity, and content validity. The researchers decided which questions were appropriate for the survey by ensuring that they were interpreted by students and teachers as intended, each dimension had multiple items, and each item corresponded with exactly one dimension. They also removed questions that required students to make inferences about other students, items with low inter-item correlations, and ambiguous questions. The study is significant because CUREs demonstrated a higher average scoring on the discovery and iteration dimensions, indicating that students in these settings have more positive outcomes because "relating science to their daily lives leads to…increased excitement and engagement" (Brownell and Kloser, 2015). Iteration is also a prominent part of the discovery process because students are not working toward previously determined findings, so repeating the process is necessary for confirming reproducibility of the results and eliminating sources of error. Collaboration, though highly scored for both CUREs and traditional labs, required more critical thinking and metacognition in research settings because students were interpreting and rationalizing their observations instead of passively sharing data. Based on the study's results, LCAS could potentially distinguish between research-based labs and traditional lab settings.

Summary

While cookbook labs teach students that science is a straightforward endeavor with little critical thinking, inquiry-based labs provide a more realistic science experience: students are more engaged, obtain a deeper understanding of concepts, and practice skills necessary for future problem-solving. Inquiry-based labs should ask students about scientific concepts used to create their data set, rather than ask to produce a desired data set. Furthermore, most labs can be converted into inquiry-based labs, by training instructors on what inquiry entails.

Introduction

This research study's broader goals are to: 1) design an observation protocol to identify inquiry in science laboratories, 2) use the data collected to compare academic learning with inquiry, 3) train TAs and head lab instructors to use inquiry practices, and 4) adjust science lab curricula to include inquiry components. To begin, this paper only focuses on the first step of this extensive research plan.

In order to obtain data, we need to settle on a set of elements that define inquiry. We must extend beyond the research of

Bruck and Fay by using a student-centered definition that "takes into account the impact of students' previous knowledge on the activity, and finds specific evidence of students' inductive reasoning, acting as an internal standard of sorts that would provide necessary controls for any assessment, and answer some of the questions regarding the existence and degree of inquiry in an activity" (Briggs, 2011). Using literature to identify trends among inquiry practices, the UTOPIA form encompasses all of these elements to determine the level of inquiry in an undergraduate lab, beyond just the extent to which students independently follow out the scientific method. After developing UTOPIA, the form was piloted in a lab course, tested for inter-rater reliability, and modified.

Methodology

UTOPIA was inspired by Project SPROUT (Simple Protocol for Observing Undergraduate Teaching) and its research team at the University of California, Irvine, funded by National Science Foundation Grant #1256500. The observers used SPROUT to identify good teaching practices, as defined by literature, in STEM lectures at UCI to connect educational practices with positive student outcomes.

UTOPIA adapted to laboratory courses instead of the lecture setting used by previous educator evaluation tools. The project adopted sections from previously validated tools such as SPROUT, the MET project, and LCAS. UTOPIA uses subdivisions from SPROUT, including Background Information and Lesson Overviews to provide environmental context and evidence throughout the observed course. It uses the concepts outlined in the MET project to develop scientifically significant data worthy of interpretation. MET project researchers advise that scoring for protocols must be rigorous, standardized, and objective, preferably with mandatory training for observers. To satisfy these guidelines, UTOPIA uses a definition of inquiry from the NRC and recommends that only "inquiry experts" act as observers, based on Espinosa-Bueno's definition of an inquiry expert. While the MET project eliminated questions that required effort beyond observation, UTOPIA does depend on carefully reading lab manuals and understanding the work expected of students before and after lab, since these sources will reveal the true extent to which inquiry is being done. UTOPIA is also designed to have content-based validity by measuring observations of the class to ensure inquiry is occurring within the lab itself.

Finally, UTOPIA was formed from studying the work of Corwin and the LCAS tool. UTOPIA does not require dimensionality because this form is based on observations instead of surveys. Repeating dimensions throughout UTOPIA would be time consuming and unnecessary; however, there are sections for evidence to be recorded for each observer's response to improve accuracy. Reliability, or consistency of results across populations, can be tested in future stages, including labs of different disciplines, lower and upper division labs, and across universities. UTOPIA incorporates the three dimensions from LCAS: collaboration, discovery, and iteration. Corwin's work found collaboration in both inquiry and traditional labs, but with varying cognitive loads. Version 4 of UTOPIA distinguishes between collaboration with and without critical thinking.

UTOPIA was piloted in Spring 2016 for 9 weeks in an introductory chemistry laboratory course. This course is a requirement for non-chemistry STEM majors, primarily those in biology and public health. The students were all freshmen who had never taken an undergraduate lab before, and the classes included 24 students in each 4-hour lab. The tenth-week class consisted of a practical exam, which was not considered for this project. Unlike most of the university's chemistry labs, this course does not include a corresponding laboratory lecture, which typically reviews or teaches the concepts needed to understand each week's lab, answers pre-lab questions, and gives experimental tips for timing and yield. The UTOPIA researcher observed the same section each week, meaning the students, teaching assistant (TA), time, and location were all consistent. To test inter-rater reliability, a colleague familiar with NGSS, inquiry, and chemistry lab subject matter co-observed the same course in Spring 2018 after being trained in how to use UTOPIA.

Results

Each lab began with the instructor reviewing the procedure and data to collect with pre-written summaries, flowcharts, or diagrams on a whiteboard. Sometimes students were asked questions that required a group response. After the introduction, students worked in pairs to complete the lab.

The intention of UTOPIA was to use conversation as evidence for the cognitive level of thinking by students throughout the lab. In reality, freshmen speak very little to each other when paired with strangers in their first undergraduate lab course. Additionally, the observed TA sat at the front desk and responded to questions only when approached. Instead of analyzing inquiry through conversation, the observation became focused on analyzing student actions to interpret thought processes. For example, if a student was observed to measure the concentration of a substance without a stir bar, then this student was assumed to lack a conceptual understanding of the chemical concentration. However, these interpretations require a strong background in the subject matter being explored, as well as an expert in inquiry practices. Because of limited interactions between the TA and students, more student-student interactions were observed. The types of comments and questions often varied based on to whom the student was talking. Questions for partners, other classmates, and the instructor had different themes, so UTOPIA was modified to differentiate these in the Version 2.

The test for inter-rater reliability revealed different interpretations for several of the questions, so these were modified for Version 3, and examples for clarification were added for many of the statements in UTOPIA. Comparisons between the two researchers' results were made for Sections III-VIII of UTOPIA, as shown in Table 3. The percentage of answers in agreement was 76%, above the accepted 70% threshold necessary for robustness.

Table 3

Results for inter-rater reliability

Section	Answers in Agreement	Answers Not in Agreement
III. Teaching Methods	4	1
IV. Discovery	3	1
V. Scientific Method	4	2
VI. Collaboration	6	1
VII. Iteration	3	2
VIII. Instructor-student Interactions	5	1
Total	**25**	**8**

Discussion

After modifying UTOPIA to distinguish the types of questions observed in Version 2, the test for inter-rater reliability (IRR) was completed. An in-depth analysis for differences in responses is given in Table 4. Four items resulted in modifying statements or adding examples to UTOPIA, Version 3. Three items did not require adjustments. The remaining two items revealed a major problem with UTOPIA that needs resolution before the next phase of this project.

Results from UTOPIA strongly depend on the observers' focus toward either the curriculum or the environment, called the curriculum effect. Potentially, there could be two versions of the protocol, depending on the situation. A curriculum-centered UTOPIA could be used to evaluate and standardize tests across the nation. An instructor-centered UTOPIA could be used by departments for internal evaluations and to create customized TA trainings based on results of their instruction.

Analyzing patterns across the nine weeks found that the level of inquiry varies based on students and TAs. When students look up the "correct" answer to a lab, they are not thinking critically about the scientific concepts behind their data. There should be a way to hold students more accountable for their work and limit variability of inquiry within a curriculum based on the effort of students. While UTOPA was piloted with a consistent TA, the test for IRR occurred outside the nine-week pilot with a different TA. A TA's style for leading the course also affected the level of inquiry. To minimize effect from TAs, UTOPIA recommends mandatory training before teaching a lab course regarding pedagogy and science education, since their backgrounds are primarily in the subject matter taught. The training should also define the expectations of a TA. All students should have the opportunity to be conceptually challenged, regardless of the TA.

Conclusion

While UTOPIA is not ready to be used for evaluating inquiry at this phase, the form has been found to be valid and robust. Once the protocol is modified for content validity, it will distinguish between measuring a curriculum and measuring an instructor's performance. The next phase includes converting UTOPIA into a tool for quantitatively evaluating inquiry on a more complex level than previously developed. Once perfected, the number of questions to be investigated is limitless. Examples include:

1. Is there more inquiry in majors-only labs or those for non-majors only?
2. Is there more inquiry in lower or upper division laboratory courses?
3. Do students in higher inquiry labs have greater retention in their STEM major?

With further development, funding, and trained staff, this tool could be used to create a standard for measuring inquiry in undergraduate labs, assess and rank programs, and inspire a movement to rebuild laboratory courses to prepare students with 21st-century skills.

Table 4

An Analysis of Disputed Statements in the IRR Study

Disputed Statement	Comments	Solution
Does the TA demonstrate use of equipment?	At the beginning of the lab, TA demonstrated use of a spill kit as a mandatory series of safety trainings at the beginning of lab. One researcher counted this as demonstrating use of equipment, but the other researcher found it irrelevant because she did not demonstrate use of any equipment needed for the day's lab.	Clarify "equipment necessary for this lab"
Students give priority to evidence in lab	One researcher commented that all post-lab questions require use of their collected data, so they are giving priority to evidence in the lab to find results. The other researcher quoted a conversation between two students who discussed whether their data was within the accepted range, using prior knowledge instead of lab evidence.	The curriculum effect is a concern for UTOPIA. Clarifying statements and training is necessary to distinguish between evaluating curriculum and observations.
Students develop a scientific explanation based on evidence	One researcher claimed the class was developing a scientific explanation (rate law) based on evidence (evaluating an equation with data). The other researcher did not consider the rate law to be a scientific explanation, but instead a phenomenon that could be explained with a scientific explanation. However, this explanation was not asked for in the post-lab questions.	Clarification is needed on what a "scientific explanation" includes
Students connect their explanation to scientific knowledge	One researcher claimed that students needed to connect their calculated rate law to the theory and background in their post-lab questions. However, their post lab exclusively asked quantitative questions, such as the rate constant or order of reaction.	No adjustments needed, human error—observers need to carefully analyze the lab manual for accuracy
Students work together for data analysis	Both researchers acknowledge that students work together for data analysis, but one researcher marked "no" for this question because it was not intended for students to work together for post-lab questions.	Curriculum effect—Clarify if observers should give priority to the intended curriculum or observed data
Students revise or repeat procedures to account for errors	Students complete the same steps twice, but the second trial is with a more dilute solution. One researcher interpreted this as iteration, and the other did not because they did not repeat the exact same procedure.	Clarification needed on the purpose and specifics of iteration
Students revise analyses based on feedback	One researcher saw students revise analyses in their Electronic Lab Notebooks after discussing their graphs with the TA	No improvements needed—sometimes observers will not witness every interaction in the room. With a large enough data set, generalizability is still acceptable
TA asks students conceptual questions	At the beginning of lab, the TA asked questions, some of which were conceptual. One researcher, however, marked "no" because one-word answers are not instinctually sufficient to answer conceptual questions. After debriefing, the researcher realized that they were conceptual and opted to change her response.	Clarify both "conceptual" and "high-level," since these can be distinguished, but not easily separated. Rephrase as "TA asks students conceptual, high cognitive level, or critical thinking questions"

Acknowledgements

I would like to thank my faculty advisor from the School of Education, Dr. Mark Warschauer, for assistance in writing this paper; graduate researcher for Project SPROUT, Dr. Lynn Reimer, for inspiring this project; and project manager of SPROUT, Amanda Nili, for helping transform my interest in STEM education into a researchable question. I would also like to thank Dr. Kimberly Edwards for her cooperation in the testing stages of UTOPIA, Tiffany Banh for her assistance with interrater reliability, the entire SPROUT team for their contributions to STEM education research that led to this project, and Said Shokair for encouraging UCI students to become passionate contributors to society.

Works Cited

Alozie, Nonye. "Promoting 21st-Century Skills in the Science Classroom by Adapting Cookbook Lab Activities: The Case of DNA Extraction of Wheat Germ." The American Biology Teacher 74.7 (2012): 485–89.JSTOR. Web. 14 June 2015.

Anderson, Ronald. "Reforming Science Teaching: What Research Says About Inquiry." Journal of Science Teacher Education 13.1 (2002): 1–12. Web. <https://link.springer.com/article/10.1023%2FA%3A1015171124982>.

Asay, L. D., & Orgill, M. (2010). Analysis of essential features of inquiry found in articles published in *The Science Teacher*, 1998–2007. Journal of Science Teacher Education, 21.

Bell, Randy et al. "Simplifying Inquiry Instruction." Nsta.Org, 2005, http://www.nsta.org/publications/news/story.aspx?id=50983.

Briggs, Michael, George Long, and Katrina Owens. "Qualitative Assessment of Inquiry-Based Teaching Methods." Journal of Chemical Education 88.8 (2011): 1034–040. American Chemical Society. Web. <http://pubs.acs.org/doi/abs/10.1021/ed100496t>.

Brown, Patrick, Sandra Abell, Abdulkadir Demir, and Francia Schmidt. "College Science Teachers' Views of Classroom Inquiry." Science Education 90.5 (2006): 784–802. John Wiley & Sons, Inc. Subscription Department, 31 Aug. 2006. Web. 14 June 2015. <https://eric.ed.gov/?id=EJ759984>.

Brownell, Sara E., and Matthew J. Kloser. "Toward a Conceptual Framework for Measuring the Effectiveness of Course-based Undergraduate Research Experiences in Undergraduate Biology." Studies in Higher Education 40.3 (2015): 525–44. Web. <http://www.tandfonline.com/doi/abs/10.1080/03075079.2015.1004234>.

Bruck, Laura, Stacey Bretz, and Marcy Towns. "Characterizing the Level of Inquiry in the Undergraduate Laboratory." Journal of College Science Teaching (2008): 52–58. Research and Teaching. National Science Teachers Association. Web. <http://www.chem.purdue.edu/towns/Towns%20Publications/Bruck%20Bretz%20Towns%202008.pdf>.

Corwin, Lisa, Christopher Runyon, Aspen Robinson, and Erin Dolan. "The Laboratory Course Assessment Survey: A Tool to Measure Three Dimensions of Research-Course Design." CBE—Life Sciences Education 14.4 (2015): 1–11. Web. <http://www.lifescied.org/content/14/4/ar37.full.pdf+html>.

Donath, Lori, et al. "Characterizing Discourse Among Undergraduate Researchers in an Inquiry-Based Community of Practice." Journal of Engineering Education, 94.4 (2005): 403–417.

Eddy, Sarah, Mercedes Converse, and Mary Wenderoth. "PORTAAL: A Classroom Observation Tool Assessing Evidence-Based Teaching Practices for Active Learning in Large Science, Technology, Engineering, and Mathematics Classes." CBE-Life Sciences Education 14 (2015): 1–16. American Society for Cell Biology. Web.

Espinosa-Bueno, Juana, Diana Labastida-Pina, Kira Padilla-Martínez, and Andoni Garritz. "Pedagogical Content Knowledge of Inquiry: An Instrument to Assess It and Its Application to High School In-Service Science Teachers." US-China Education Review 8.5 (2011): 599–614. David Publishing. Web. <http://files.eric.ed.gov/fulltext/ED520684.pdf>.

Fay, Michael, Nathaniel Grove, Marcy Towns, and Stacey Bretz. "A Rubric to Characterize Inquiry in the Undergraduate Chemistry Laboratory." Chemistry Education Research and Practice 8.2 (2007): 212–19. Royal Society of Chemistry. Web.

Holding, Matthew. "Confronting Scientific Misconceptions by Fostering a Classroom of Scientists in the Introductory Biology Lab." The American Biology Teacher 76.8 (2014): 518–23. JSTOR. Web.

Joe, Jilliam, Cynthia Tocci, and Steven Holtzman. MET Project Policy and Practice Brief. Rep. N.p.: n.p., 2013. Foundations of Observation: Considerations for Developing a Classroom Observation System That Helps Districts Achieve Constant and Accurate Scores. Bill and Melinda Gates Foundation. Web. <http://k12education.gatesfoundation.org/download/?Num=2671&filename=Classroom_Observation_092110.pdf>.

Kirschner, Paul A. "Epistemology, Practical Work and Academic Skills in Science Education." Science and Education, 1. 3 (1992): 273–299.

Kirschner, Paul A., and M. A. M. Meester. "The Laboratory in Higher Science Education: Problems, Premises and Objectives." High Educ Higher Education 17.1 (1988): 81–98.

Marshall, J. C., Horton, B., Smart, J., & Llewellyn, D. (2008). EQUIP: Electronic Quality of Inquiry Protocol: Retrieved from Clemson University's Inquiry in Motion Midwest Comprehensive Center at the American Institutes for Research (AIR) & the Wisconsin Department of Public Instruction (WDPI). 2016. What ever happened to scientific inquiry? A look at evolving notions of inquiry within the science education community and national standards. 7–8.

Mohrig, Jerry R. "The Problem with Organic Chemistry Labs." Journal of Chemical Education, 81. 8, (2004): 1083

National Research Council (NRC). 1996. National science education standards. Washington, DC: National Academy Press.

National Research Council (NRC). 2000. Inquiry and the national science education standards: A guide for teaching and learning. Washington, DC: National Academy Press.

Saxton, Emily, Robin Burns, Susan Holveck, and Sybil Kelley. "A Common Measurement System for K-12 STEM Education: Adopting an Educational Evaluation Methodology That Elevates Theoretical Foundations and Systems Thinking." Studies in Educational Evaluation 40 (2014): 18–35. Science Direct. Elsevier. Web. <http://www.sciencedirect.com/science/article/pii/S0191491X13000679>.

Timmerman, B. E., D. C. Strickland, and S. M. Carstensen. "Curricular Reform and Inquiry Teaching in Biology: Where Are Our Efforts Most Fruitfully Invested?" Integrative and Comparative Biology 48.2 (2008): 226–40. JSTOR. Web.

"Transitioning from Scientific Inquiry to Three-Dimensional Teaching and Learning." NSTA Position Statement, NSTA Board of Directors, Feb. 2018, www.nsta.org/about/positions/3d.aspx.

Unsworth, Elizabeth. "Macromolecules Inquiry: Transformation of a Standard Biochemistry Lab." The American Biology Teacher 76.7 (2014): 438–41. JSTOR. Web.

Velasco, Jonathan B., et al. "Characterizing Instructional Practices in the Laboratory: The Laboratory Observation Protocol for Undergraduate STEM." Journal of Chemical Education, vol. 93, no. 7, 2016, pp. 1191–1203., doi:10.1021/acs.jchemed.6b00062.