Biological Sciences
where discovery comes to life UC San Diego

Assessment of scientific reasoning skills learned from lab modules with varying degrees of inquiry

Aaron Coleman, Sahil Khan, Goran Bozinovic Division of Biological Sciences, University of California, San Diego

Introduction

In addition to learning the facts and concepts of biology, we hope that our students are learning how science works as a process. This includes the reasoning skills required to form hypotheses, analyze data, and design experiments to test a given question. These skills are best learned in laboratory classes, where students can engage in the scientific process. Recent calls for reform in biology undergraduate education have urged a change to more inquiry-based laboratory classes. It is logically assumed that students should learn scientific reasoning better from inquiry-based classes. Participation in thinking about what the experimental questions should be to study a particular topic and then reasoning how to design an appropriate experiment are thought to lead to better student outcomes, as opposed to conducting "cookbook"-style experiments to replicate predetermined results (Handelsman *et al.*, 2004). While some studies have shown better learning of scientific reasoning from inquiry-based labs (Lord and Orkwiszewski, 2006), evidence from controlled studies on more advanced, upper-division classes, is sparse.

A recent meta-analysis of 142 papers describing inquiry-based labs found that the definition of inquiry varies widely (Beck, *et al.*, 2014). Inquiry can range from the students determining the answer to a question (structured inquiry), to the students defining the hypothesis and all aspects of the experiment (open-inquiry)(Windschitl, 2002). This begs the question, how much inquiry is necessary to improve student outcomes? From the practical standpoint of designing labs for high-enrollment courses, greater inquiry poses more logistical challenges and incurs greater expense.

We have designed an assessment to measure scientific reasoning skills, and administered it to students before and after completion of our upper-division biochemistry lab course. The assessment was used to compare outcomes from low-inquiry and high-inquiry versions of the course.

Course Studied and Inquiry Rubric

Biochemical Techniques is an upper-division biochemistry and molecular biology lab course. We routinely rotate two different versions of the biggest project through the class in different quarters. In both versions of the project the students examine signal transduction and conduct a Western blot to measure Erk/MAP kinase phosphorylation (activation), and an ELISA to measure phospholipase C activation.

Version 1—Low Inquiry: This version of the project would be defined as structured inquiry. The experiments are applied to study sea urchin fertilization. The students are provided with the questions that they will answer and the experimental variables they will perform. Over the course of the project they answer the questions "Is the calcium influx that occurs at fertilization sufficient for MAP kinase inactivation, and is the calcium influx sufficient for cleavage of the zygote?"

Version 2—High inquiry: This version of the project would be somewhere between guided and open inquiry. The students study fibroblast growth factor (FGF-2) signal transduction in cultures of NIH 3T3 fibroblasts. They are provided with a dataset that describes three effects of FGF-2 in NIH 3T3 cells, and based on this they must develop a hypothesis to explain how these effects are signaled and design an experiment to test it.

	Inquiry Element	Sea urchin signaling—Low Inquiry	FGF signaling—High-Inquiry
1	Background concepts, information on experimental system	provided	provided
2	Make observations about system, interpret data	not done	derive
3	Derive question or hypothesis	provided	derive
4	Procedures, techniques used	provided	provided
5	Design experiment, come up with variables	provided	derive
6	Decide how to analyze raw data	provided	provided
7	Derive conclusions from data	derive	derive
8	Communicate results	provided	provided

Rubric of Inquiry Elements We have developed a rubric for defining the degree of inquiry in a laboratory project that is suitable for upper-division biology labs, revised from those developed by Buck *et al.* (2008) and Weaver *et al.* (2008). The table gives each of the eight rubric items and indicates for the two versions of the BIBC 103 lab project whether that item is **provided** to the students, or if the students **derive** that item, or if it is **not done**.

The Scientific Reasoning Skills Assessment

We created an assessment to measure scientific reasoning skills that would meet the following requirements:

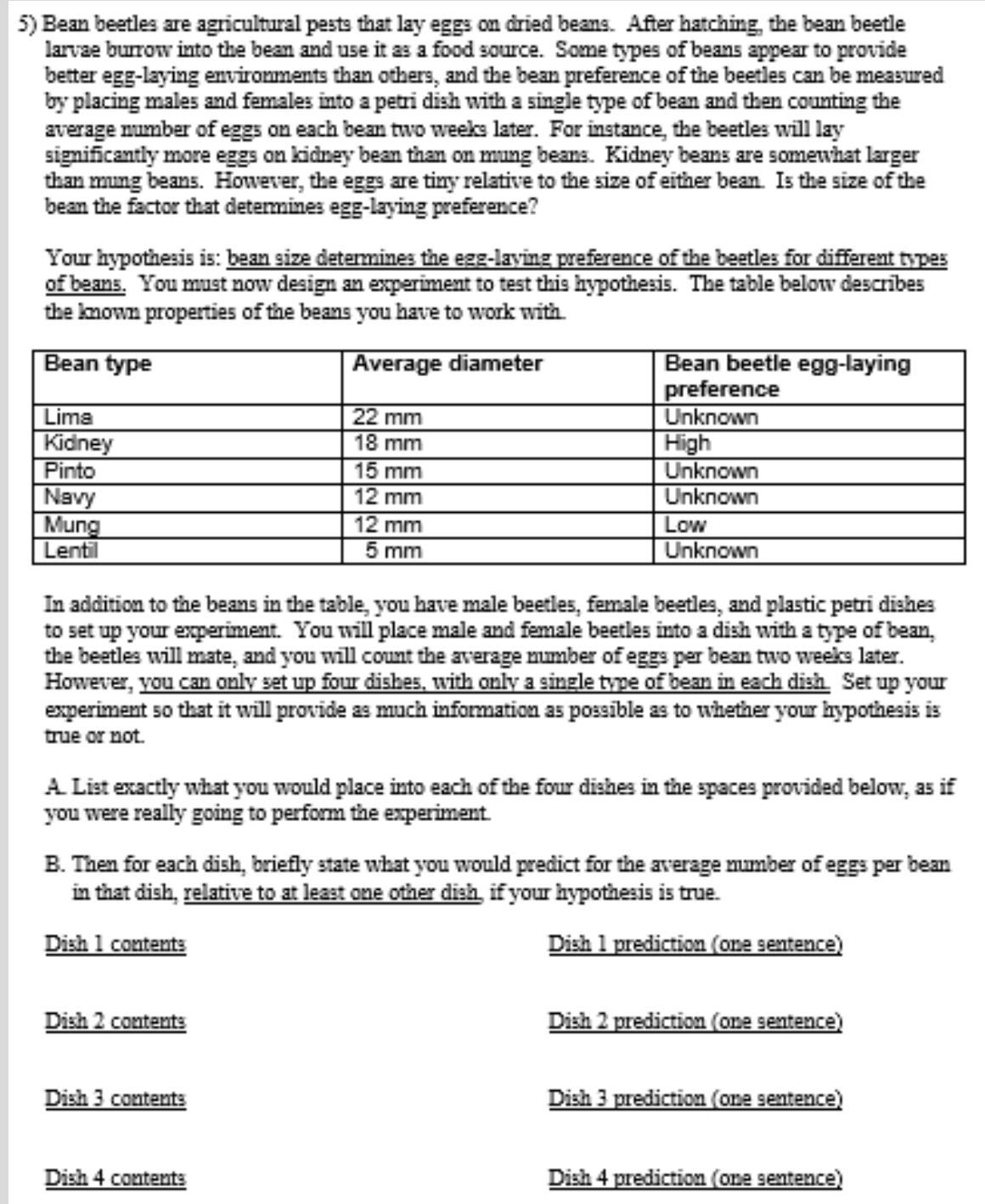
- **Questions of varying difficulty to capture a range of skill levels**
- **Can be completed in relatively short period of time (45 minutes)**
- **Scoring is not too time consuming**
- **Clear rubric for free-response questions**

Assessment Design

Skill Assessed	Question	Type	Point value	Relative difficulty
Form Hypothesis	1	free response	2	easy
	2	free response	2	difficult
Interpret Data	3	multiple choice	2	easy – moderate
	4	multiple choice	2	difficult
Design Experiment	5	free response	4	varied – multiple components

Assessment Validation

Six students were interviewed after completing the post class assessment and asked about each question. During the interview, the complete answer to each question was carefully explained to the student. When asked about how understandable each question was with regard to what we intended, the students generally responded that the questions were very clear. The students were also asked if with sufficient instruction in the particular skill tested by that question, whether it was reasonable that it could be answered by an undergraduate student in biology. For free response questions that were scored by looking for specific rubric items to be mentioned (see sample assessment question), the students were asked about each rubric item individually. Students generally responded that it was very likely each question could be answered correctly by an undergraduate. One exception was one of the rubric items looked for in free response answers to question 2 (see the Assessment Design table). Four of the six students interviewed responded that it was not very likely an undergraduate student could generate that response, one responded it was somewhat likely and one responded it was very likely. A small number of students (0.5 – 1%) do give the complete full answer to this question when taking the assessment.


5) Bean beetles are agricultural pests that lay eggs on dried beans. After hatching, the bean beetle larvae burrow into the bean and use it as a food source. Some types of beans appear to provide better egg-laying environments than others, and the bean preference of the beetles can be measured by placing males and females into a petri dish with a single type of bean and then counting the average number of eggs on each bean two weeks later. For instance, the beetles will lay significantly more eggs on kidney bean than on mung beans. Kidney beans are somewhat larger than mung beans. However, the eggs are tiny relative to the size of either bean. Is the size of the bean the factor that determines egg-laying preference?

Your hypothesis is: bean size determines the egg-laying preference of the beetles for different types of beans. You must now design an experiment to test this hypothesis. The table below describes the known properties of the beans you have to work with.

Bean type	Average diameter	Bean beetle egg-laying preference
Lima	22 mm	Unknown
Kidney	18 mm	High
Pinto	15 mm	Unknown
Navy	12 mm	Unknown
Mung	12 mm	Low
Lentil	5 mm	Unknown

In addition to the beans in the table, you have male beetles, female beetles, and plastic petri dishes to set up your experiment. You will place male and female beetles into a dish with a type of bean, the beetles will mate, and you will count the average number of eggs per bean two weeks later. However, you can only set up four dishes, with only a single type of bean in each dish. Set up your experiment so that it will provide as much information as possible as to whether your hypothesis is true or not.

A. List exactly what you would place into each of the four dishes in the spaces provided below, as if you were really going to perform the experiment.

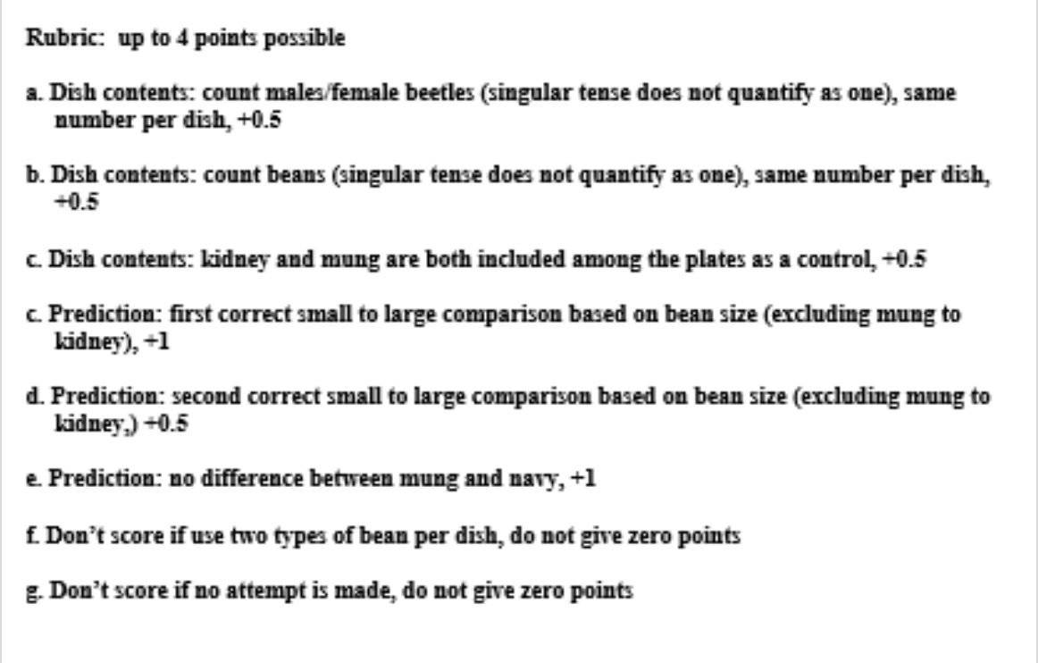
B. Then for each dish, briefly state what you would predict for the average number of eggs per bean in that dish, relative to at least one other dish, if your hypothesis is true.

Dish 1 contents Dish 1 prediction (one sentence)

Dish 2 contents Dish 2 prediction (one sentence)

Dish 3 contents Dish 3 prediction (one sentence)

Dish 4 contents Dish 4 prediction (one sentence)


Rubric: up to 4 points possible

a. Dish contents: count males/female beetles (singular tense does not quantify as one), same number per dish, +0.5

b. Dish contents: count beans (singular tense does not quantify as one), same number per dish, +0.5

c. Dish contents: kidney and mung are both included among the plates as a control, +0.5

c. Prediction: first correct small to large comparison based on bean size (excluding mung to kidney), +1

d. Prediction: second correct small to large comparison based on bean size (excluding mung to kidney,) +0.5

e. Prediction: no difference between mung and navy, +1

f. Don't score if use two types of bean per dish, do not give zero points

g. Don't score if no attempt is made, do not give zero points

Example Question from Assessment Question 5 from the assessment is shown above, along with the rubric for scoring the question.

Results

	Instructor A Fall 2017	Instructor A Spring 2018	Instructor B Winter 2018	Instructor B Winter 2019	Instructor C Winter 2019
Degree of inquiry	High inquiry	Low inquiry	High inquiry	Low inquiry	Low inquiry
Number of students	77	68	50	106	21
Mean score on assessment (out of 12 points)	Pre 6.27 Post 6.82	Pre 6.65 Post 6.69	Pre 6.61 Post 7.23	Pre 5.48 Post 6.20	Pre 6.07 Post 6.38
$p =$	< 0.01	not significant	< 0.05	< 0.001	not significant
Effect size (Cohen's d)	0.30 (small effect)	0.03 (no effect)	0.36 (small effect)	0.42 (small/med. effect)	0.18 (no/small effect)

Assessment Results: The assessment was administered at the beginning and end of the class, for five independent classes taught by three different instructors. The students were incentivized by offering a small amount of extra credit, based on the number of question answered correctly, for instructors A and C. For instructor B, the students were not incentivized. The pre and post-test assessments results for each student were paired for the analysis.

Conclusions and Discussion

The assessment: The scientific reasoning skills assessment was able to measure pre to post-class gains in these skills across multiple classes taught by different instructors. Three of the 5 classes assessed showed a statistically significant increase for the average score. Along with the positive response to the assessment questions by students who were interviewed after completing it, this suggests the assessment is a valid instrument for measuring learning gains in the skills required for hypothesis formation, data interpretation, and experimental design. While the change in average pre to post-class assessments scores was modest, this is not unexpected. The more difficult assessment questions were designed to be challenging to students who had completed the Biochemical Techniques course so that a broad range of skill levels could be measured. This is in contrast to what would be expected for pre to post-class improvement on a concept inventory, where students who have successfully learned the material are expected to be correctly answer the majority of the questions.

The high-inquiry version of the class may achieve more consistent improvement in scientific reasoning skills: The two high-inquiry classes showed statistically significant increases in student learning of scientific reasoning skills, with a small effect size as measured by Cohen's *d*. Two of the three low-inquiry classes showed no improvement or non-significant improvement in average score. However, the third low inquiry-class showed strong, significant improvement with a close to moderate effect size. While improvement in scientific reasoning skills may be more consistent with the high-inquiry version of the class, it is possible that students may learn these skills equally well in the low-inquiry version. The low-inquiry class project does have some elements of inquiry, and involves more than simply replicating known results.

Learning scientific reasoning skills is only one of the student outcomes that we hope to achieve in biology lab classes. In addition to mastery of important biological concepts, other outcomes include improvement in science self-efficacy and self-identity, and a more positive attitude toward learning science. Inquiry-based labs have been shown to improve student outcomes in these areas that involve student attitudes (Beck *et al.*, 2014; Lord and Orkwiszewski, 2006). In particular, student feelings of ownership toward their work in biology lab classes are improved with the introduction of inquiry-based curriculum (Hester *et al.*, 2018). In the next phase of our study we plan to use the Project Ownership Survey (Hanauer and Dolan, 2014) to measure differences in this affect between students who have completed the low vs. high-inquiry versions of the class.

References

Beck, C. *et al.* 2014. Promoting inquiry-based teaching in laboratory courses: Are we making the grade? *CBE-Life Sciences Education* 13:444-452.
Buck, L.B. *et al.* 2008. Characterizing the level of inquiry in the undergraduate laboratory. *Journal of College Science Teaching* 38:52-58.
Hanauer, D.I. and Dolan, E.L. 2014. The project ownership survey: Measuring differences in scientific inquiry experiences. *CBE-Life Sciences Education* 13:149-158.
Handelsman, J. *et al.* 2004. Scientific teaching. *Science* 304:521-522.
Hester, S.D. *et al.* 2018. Authentic inquiry through modeling in biology (AIM-Bio): An introductory laboratory curriculum that increases undergraduates' scientific agency and skills. *CBE-Life Sciences Education* 17:1-23.
Lord, T. and Orkwiszewski, T. 2006. Moving from didactic to inquiry-based instruction in a science laboratory. *The American Biology Teacher* 68:342-345.
Weaver, G.C. *et al.* 2008. Inquiry-based and research-based laboratory pedagogies in undergraduate science. *Nature Chemical Biology* 4:577-580.
Windschitl, M. 2002. Inquiry projects in science teacher education: What can investigative experiences reveal about teacher thinking and eventual classroom practice? *Science Education* 87:112-143.