

Comparing Physical, Virtual, and Hybrid Flipped Labs for General Education Biology

Ji Y. Son

Department of Psychology, California State University, Los Angeles

Paul Narguizian, Dwight Beltz, and Robert A. Desharnais

Department of Biology, California State University, Los Angeles

Authors Note

This research was supported by grants from the California State University (Course Redesign with Technology Initiative) and U.S. National Science Foundation (DMS-1225529). Correspondence concerning this article should be addressed to Ji Y. Son of Psychology, California State University, 5151 State University Drive, Los Angeles, CA 90032. Email: json2@calstatela.edu

Abstract

The purpose of this study was to examine the impact on learning, attitudes, and costs in a redesigned general education undergraduate biology course that implemented web-based virtual labs (VLs) to replace traditional physical labs (PLs). Over an academic year, two new modes of VL instruction were compared to the traditional PL offering: (1) all VL with an in-person help center (VL-A) and (2) a hybrid flipped VL model where online labs alternated with in-person labs every week (VL-H). All three lab types included a face-to-face lecture with the same materials. Engaging inquiry-based exercises were developed for each VL activity in which students were provided background information, guided through a series of basic experiments, encouraged to design their own experiments, and required to produce a simple scientific report that was delivered for evaluation electronically. The VL-A group had the highest proportion of repeatable grades (below a C, 2.0 grade points). Students in the VL-H group achieved significantly better grades compared to the other lab instruction groups. The VL-H group also experienced statistically significant favorable shifts in their self-reported attitudes towards biology. The personnel costs for the VL-A and VL-H models were 29% and 63% of the PL model, respectively, allowing more sections to be offered. These results suggest that carefully designed online lab opportunities can result in higher student grades and more favorable attitudes towards science while reducing costs compared to traditional labs.

Introduction

General education (GE) science requirements are designed to provide students with intellectual tools for growth and a broad base of knowledge about the natural sciences. Part of that mission is fulfilled by the inclusion of laboratory work generally assumed to provide experience with the process and methods of science (National Research Council, 2006). Laboratory work is supposed to simulate scientific inquiry, broadly defined as the way scientists study the natural world, propose ideas, justify assertions, and derive explanations based on evidence. Given the goals of GE science, some have argued that the principal focus of laboratory activities should not be devoted to mastery of particular laboratory techniques (see Hodson, 1993). Instead, the lab component should encourage students to investigate phenomena, solve problems, and pursue inquiry and interests in science. It follows from these assumptions that any investigation regarding laboratory pedagogy should measure students' learning as well as their attitudes towards science. The purpose of this study was to compare learning and attitude outcomes alongside cost considerations in three different implementations of the lab component of a GE biology course: a physical lab (PL), an all virtual lab (VL-A), and a hybrid "flipped" lab (VL-H). Can a GE science course designed with a virtual lab (VL) experience fulfill the vision of encouraging students to pursue inquiry and interest in science?

Part of the motivation for this study comes from the practical problem of offering high quality PL experience to every student. At Cal State LA, as is commonly the case in many institutions of higher education, a science course with a laboratory activity is required of every student. For practical reasons, the enrollment in each lab section must be limited in size (20-24 students) and staffed by an instructor or graduate teaching assistant. Lab sections must be taught in specialized facilities that are limited in number and availability. VLS could potentially address this resource bottleneck by removing some of the barriers to enrollments, such as the availability of laboratory space and instructors, allowing more students to be served.

Moreover, many GE students struggle with science classes and so the rates of repeatable grades are often high. Thus GE laboratory courses represent a pedagogical bottleneck for students' progress to graduation. Given that many students come from K-12 education with a less than adequate conception of the nature of science and often view science as a static body of facts (Lederman, 1992; Welch, Klopfer, Aikenhead, & Robinson, 1981), GE science courses that emphasize the process of scientific inquiry pose a particular challenge for non-science majors. Although labs are intended to be the solution to this problem by involving students in the scientific process, there are not enough resources to allow students to explore phenomena and try a variety of experiments. Even though laboratory work has great potential to provide learners with opportunities to manipulate materials and construct their knowledge of phenomena and related scientific concepts (Linn et al., 2006), some have questioned whether these are implemented widely given the lack of evidence that such critical thinking opportunities were offered in school versions of many lab activities (Roth, 1994; Tobin, 1990). Most PLs (also called "wet labs" in biology) are "cookbook" activities in which students follow specific directions and take measurements (National Research Council, 2006, 2007). These formulaic activities offer limited opportunities for creativity, which may be one reason why some students perform poorly and are less engaged in these courses.

There are undoubtedly characteristics of PLs that cannot be imitated in a virtual environment, such as experience working with specialized equipment, troubleshooting machinery, and engaging in careful set-up of studies. Even so, VLS offer two advantages to novice science learners. The first advantage is that VLS provide relatively risk-free environments for students to explore scientific concepts in inquiry-based fashion (Zacharia, Olympiou, & Papaevripidou, 2008). Using VLS, students can formulate hypotheses and carry out experiments. Mistakes are of no consequence, since modified

experiments can be redesigned with little additional effort. Students can also design and carry out more experiments and gather more information relative to the PL version of the same experiment (Klahr, Triona, & Williams, 2007). These efficiencies tend to emphasize the higher order skills required to plan experiments and to appreciate the scientific method (de Jong, 2006; Wieman, Adams, & Perkins, 2008).

The second advantage of VLs is that reality can be augmented in the service of pedagogy. Especially for novice learners, those that are new to a domain and may be easily influenced by irrelevant information, highlighting important features and stripping out unnecessary details can help direct their learning (Finkelstein et al., 2005; Goldstone & Son, 2005). For instance, novices are more likely to be misled by the noise in PLs, such as the slight displacement of an internal organ in a dissection lab or a worn battery in an electrical circuit. The idealized models present in VLs can help focus their attention on the relevant relationships between variables. Furthermore, virtual simulations can make invisible phenomena visible (e.g., depicting movements of electrons) or link observable phenomena with symbolic representations (e.g., show how variables such as heat or kinetic energy change with a reaction) (Finkelstein et al., 2005; Jacobsen & Wilensky, 2006).

Unfortunately, many of the advantages of VLs have either been theoretically proposed or narrowly tested in an experimental setting. The most rigorous experiments have compared physical and virtual versions of a single laboratory activity with a tight focus on a particular concept. Some of these experiments have shown that there are no significant differences between learning from PL and VL versions (Klahr, Triona, & Williams, 2007; Triona & Klahr, 2003; Zacharia, 2007) although students perceive PLs to be more effective than VLs (Stuckey-Mickell & Stuckey-Danner, 2007). Also, a growing body of research indicates that both PLs and VLs are most successful when students are provided sufficient guidance and opportunities for reflection (Linn et al., 2006; de Jong, 2006, 2010; Windschitl & Andre, 1998). A number of studies have also gone beyond testing PLs and VLs against each other; combining components of PLs and VLs often promotes better learning than each modality on its own (Bellido, Martinez-Jimenez, Pontes-Pedrajas & Polo, 2003; Olympiou & Zacharia, 2012; Zacharia & de Jong, 2014).

Missing from the literature are investigations comparing large-scale implementations where an entire course is comprised of PLs or VLs (Stuckey-Mickell & Stuckey-Danner, 2007, a notable exception, focused on students' perception of learning). Also there is little research examining how to integrate VLs through an entire course and provide adequate guidance for students. One reason for the lack of research on course-wide implementations is that PL and VL assignments are typically designed to emphasize their own affordances and minimize their limitations. Thus, courses in these experimental settings have wholly different sets of laboratory formats, topics, and exercises, meaning that setting up accurate comparisons would be difficult. Also, it is impossible to randomly assign students to courses because of practical limitations such as space availability (e.g., large lecture halls, laboratories) and scheduling. Despite the problems in methodology (e.g., incomparability of experiences, inability to randomize), comparisons of learning and attitude change from PL and VL courses are necessary for departments and administrators to decide whether to adopt VLs on a large scale. This kind of large-scale implementation research should be viewed in conjunction with well-controlled laboratory examinations of differences between individual PLs and VLs.

Our research program compared a PL course and two different VL courses, each designed with experiments/assignments that take advantage of their own contexts. Because the assignments, skills practiced, and, in some cases, subject matter differed as a result of the format of the labs, the goal was not to control for specific content learning. Instead, the goal of this research was to compare PLs to VLs in terms of their ability to promote inquiry-based learning and foster positive attitudes towards science at our institution. Another goal was to compare two different methods of providing student guidance in learning from VLs. The final goal was to compare the financial commitments of PLs and VLs.

To achieve these goals, we redesigned a general education biology course by using existing web-based software to replace traditional wet labs. Two new modes of VL instruction were compared to the traditional PL offering: (1) all online VL with an in-person help center (VL-A) and (2) a hybrid flipped VL model with two tracks of online VL and in-person labs alternating every week (VL-H). In VL-A and VL-H, engaging inquiry-based exercises were developed around each VL activity where students were provided background information, guided through a series of basic experiments, encouraged to design their own experiments, and required to produce a simple scientific report that was delivered electronically. The second mode of VL is considered *hybrid* because students only attended in-person every other week; however, they worked on VL group assignments (not PL assignments) when they did meet together face-to-face. The *flipped* label refers to the idea that students complete guided VL experiments first on their own and then, in a supervised group setting, design and carry out their own experiments. So the hybrid flipped VL model incorporated alternating weeks, students doing virtual experiments on their own in one week, and the next week attending an in-person lab where, under the guidance of an instructor, student discussed their results with peers, and, as a group project, planned, implemented, and analyzed their own experiments. It should be noted that the VL-H group was designed to cover fewer topics in more depth than the VL-A group because the VL-H group had an opportunity to conduct experiments alone and in a group setting. The VL-A group conducted more experiments on their own and covered more topics.

The PL differed from the VL-A and VL-H in that it followed a traditional wet lab format. Each week, students worked in small groups that physically met together. They followed instructions from a written lab manual (available online) and manipulated materials provided for each group. They recorded their observations and answered questions, turning in their lab assignment at the end of the period. There was one experiment (chronobiology) where students shared data collected at home over the course of a week and wrote a scientific report, turning in section drafts over the course of the term that were returned to the students with comments.

In all three laboratory modes (PL, VL-A, and VL-H), students engaged with questions that would typically be covered in a traditional scientific report (design, methods, results, discussion). However, PL assignments never required students to design their own experiments; instead they followed directions for a prescribed experiment. VL-A and VL-H included exercises that promoted designing experiments to address a particular issue. Also, for the PL and VL-A modes, the laboratory exercise often followed a related lecture (within a week or two). For example, after a lecture on the circulatory system, students would be assigned a lab about the cardiovascular system within one week. Because the VL-H mode had fewer topics which were covered in more depth and there were two asynchronous groups of lab sections, there was sometimes a 2-3 week delay between the lecture and the related lab topic; however, the lecture and VL-H lab topics were presented in the same order.

All versions of this GE course included a face-to-face lecture covering the same course material. These lectures were taught by three different instructors (one instructor taught the PL version, one instructor taught both VL-A and VL-H versions, and one instructor taught the VL-H version) but used the same PowerPoint slides. Although it would have been ideal to have the same instructor teaching all three modes, this was not possible because of the logistics of scheduling and hiring. Instead, the different instructors conferred regularly and used the same materials during lecture, but deviations due to style, temperament, and so forth, were not controlled. In all three versions of the course (PL, VL-A, and VL-H) there was no textbook. Instead relevant readings were provided online before the lecture. These readings were the same across courses. We verified that the patterns of data reported in this manuscript were consistent with a narrower comparison of the VL-H and VL-A versions of the course taught by the same instructor.

Students' attitudes and conceptual learning were assessed online both pre- and post-instruction. We also analyzed student achievement through grades, enrollment, and passing rates. Finally, we also examined the fiscal impact of VL implementation.

Method

Participants

Cal State LA has one of the most diverse student populations in the nation. In 2013, the students were 55.8% Hispanic, 16% Asian American, 9.9% White, and 4.7% African American. Among undergraduates, 59% of the students are female. Many students are older and have families; the average undergraduate is 23.4 years of age. Because the course was a general education course, student enrollment reflected this diversity. Three versions of this course were offered in AY 2013-14: PL model (N = 186, one section), VL-A model (N = 186, one section), and VL-H model (N = 376, two sections).

Procedure and Materials

Course background. The re-designed course was a non-majors GE science course at Cal State LA called *Animal Biology* (BIOL 155). This four-unit, quarter-based course is one of only three courses that satisfy the GE requirement for a life science course with a laboratory component. Most science majors require a non-GE biology course as part of their programs, so this course is usually taken by non-science majors. There are no prerequisites for the course. It is normally taught with two 75-minute lectures and one 150-minute laboratory session per week. The Department of Biological Sciences at Cal State LA usually offers this course 2-3 times per year with 6-8 PL sections of 24 students each in specialized laboratory facilities. The lecture is always offered in a large lecture hall (144-192 students total). The lectures are given by a tenured/tenure track or adjunct faculty member and the PL sections are most often staffed by adjunct faculty or graduate teaching assistants. Tenure/tenure track faculty may occasionally teach a few of the lab sections.

Cal State LA is on the quarter schedule so the PL and VL-A models were offered in Winter quarter (one section each) and the VL-H model (two sections) were offered in the Spring quarter. These sections were taught by experienced faculty members who had previously taught this course. The PL section was taught by an instructor who had previously taught the course with PLs only. The students in the VL-A section and one of the VL-H sections (N = 184) were taught by a different instructor who had taught PL and VL-A versions of the course. The remaining VL-H section (N = 192) was taught by an instructor who had previous experience teaching the course with PLs only. All of the lecture portions of the three models used the same syllabus and PowerPoint slides. The main differences between the offerings were the format and content of the lab sections.

PL model. In the traditional PL mode of the course, students met in person weekly to conduct exercises as outlined in a laboratory manual delivered as handouts available online. Most of these exercises required the completion of a laboratory exercise that was turned in at the end of the lab period. For example, in the exercise on digestion, students followed instructions to combine egg white (protein), canned milk (fat), and potato (starch) with various enzymes or distilled water (control) and observe what happens. Students worked in groups for every lab exercise. They also answered questions and turned their paper-based answer sheets for grading. Students were also required to submit one longer laboratory report in the format of a scientific paper with drafts due over the course of the quarter. Students generally do not have the opportunity to formulate hypotheses and design experiments on their own in these exercises. Summaries of all PL assignments are provided in Table 1.

VL-A model. In this model, all labs were offered online (summarized in Table 1). Nine VLs were employed, one for each week of the academic quarter excluding the first week when labs do not meet. Students were provided with a handout, delivered through the online course management system, with an introduction on how to use the activity and step-by-step instructions that led students through a series of experiments or activities designed around important concepts from the course. Students also answered multiple-choice questions about the VL assignment in the online course management system. For three of the nine labs, students were given an additional assignment where a problem was posed and then students had to propose a formal hypothesis, design experiments to test the hypothesis, carry out the experiments, analyze the results, and present all of this information in the format of a brief scientific report. A report template was provided with instructions for each of the following sections: introduction, experimental design, results, discussion and conclusions. A grading rubric was provided to the students to guide them in the writing of their reports.

Graduate assistants staffed a drop-in center for students with questions about the lab assignments. The help center was open for 30 hours/week. The number of student visits to the help center varied from approximately 25-100 per week, with the highest number of visits occurring at the beginning of the academic term. Questions divided equally between technology-related problems (e.g. “how do I get this applet to run on my laptop?”) and content-related inquiries (e.g. “how do I design a dihybrid genetic cross?”).

VL-H model. In this model, students met in the physical laboratory every other week. On days they met in the laboratory, students worked on group exercises under the guidance of a laboratory instructor. During alternating weeks when they were not meeting in person, each student worked on individual VL assignments. The lecture class was divided into two tracks. Students in track A met in the laboratory on even weeks of the quarter and worked on individual assignments on odd weeks. Students in track B met in the laboratory on odd weeks of the quarter and worked on individual assignments on even weeks. This process is illustrated in Table 1.

Table 1: *The Alternating Structure of the Hybrid Flipped VL Model*

Week	Track A	Track B
1	Intro to virtual lab	No lab
2	Individual online exercises	Intro to virtual lab
3	Group report in lab	Individual online exercises
4	Individual online exercises	Group report in lab
5	Group report in lab	Individual online exercises
6	Individual online exercises	Group report in lab
7	Group report in lab	Individual online exercises
8	Individual online exercises	Group report in lab
9	Group report in lab	Individual online exercises
10	No lab	Group report in lab

Lab instructors alternated face-to-face meetings with students from Tracks A and B. Thus, the same laboratory facility and number of instructors accommodated twice as many students, addressing some of the resource limitations at the institution.

During the first in-person meeting, the lab instructor explained the organization of the labs and introduced the first virtual lab activity. The following week, students individually worked on a set of step-by-step instructions that led them through a series of experiments or activities designed around important concepts from the course. These were a subset of the VL-A exercises (see Table 1). Like students in the VL-A group, VL-H students also answered multiple-choice questions online. When they came together in the next face-to-face meeting, students discussed their answers to the online exercises and then were

given a second attempt to answer the multiple choice questions. The highest grade on the two attempts counted towards their course grade. Additionally, students worked together to formulate hypotheses, design and carry out experiments to test their hypotheses, organize their results, and submit a report in the format of a scientific paper. The lab instructor also introduced the next activity during these meetings. This pattern of individual online activities followed by in-person group work was repeated until the end of the quarter.

VL Assignments. Nine virtual labs were employed during the VL-A model (see Table 1). Six were from *Biology Labs OnLine* (BLOL) and the remaining three were from *SmartScience Labs* (SSL). BLOL are simulations of experimental situations such as the genetics of inheritance or evolution. Students can vary several inputs in order to design a large variety of experiments. Tabular and graphical outputs were provided as well as the ability to transfer and export data their experimental data. SSL provides videos of real experiments that the students can view and pause to collect data. Videos of experiments conducted under different conditions are provided. The software has integrated introductory information.

A subset of four labs (all Biology Labs Online) was employed for the VL-H model. These labs were chosen because they offered more flexibility in terms of designing experiments. As part of the in-person activities, students were required to formulate hypotheses and design and carry out experiments to test their hypotheses. Table 1 describes the virtual labs that were employed in the VL-A and VL-H models. Copies of the lab handouts are available at <http://tinyurl.com/vlab-eport>.

Table 2: *Lab Activities in All Virtual (VL-A), Hybrid Virtual (VL-H), and Physical Labs (PL)*

Activity Topic	Description	Usage
<i>Biology Labs OnLine</i>		
Cardio lab	Addressed homeostasis using arterial blood pressure as an example. The interaction of variables related to heart rate, vessel radius, blood viscosity, and stroke volume are examined.	VL-A
Demography lab	Investigated differences in population size, age-structure, and age-specific fertility and mortality rates affect human population growth.	VL-A
Evolution lab	Modeled adaptation by natural selection by manipulating various parameters of a bird species and its habitat, such as initial mean beak size, variability, heritability, population size, precipitation and island size.	VL-A, VL-H
Fly lab	Taught genetic inheritance by designing mating between female and male fruit flies carrying one or more genetic mutations.	VL-A, VL-H
PopEco lab	Provided an example of population ecology by manipulating the life history attributes of three bird species: two competing sparrows and a hawk predator.	VL-A, VL-H
Translation lab	Featured characteristics of the genetic code by creating and translating simple RNA sequences into polypeptides.	VL-A, VL-H
<i>Smart Science Labs</i>		
Animal behavior	Profiled behavior of pill bugs in different situations to determine how these particular animals respond to their environment.	VL-A
Enzymes & pH	Harnessed effect of pH on enzyme reactions. They explore how the rate of reaction changes as a function of pH.	VL-A
Frog dissection	Provided videos of a frog dissection and learn to identify different organs and their functions.	VL-A
<i>Physical Labs</i>		
Scientific method	Overviewed the scientific method to investigate the length of appendages versus total body length in humans.	PL

Comparing Physical, Virtual, and Hybrid Flipped Labs for General Education Biology

Chronobiology	Captured student circadian rhythms at specified times throughout the day by measuring their pulse rate, eye-hand coordination and adding speed. Students pool their data and write a scientific report.	PL
Movement through membranes	Assigned four simple experiments that demonstrated diffusion, dialysis, osmosis, and plasmolysis.	PL
Digestion	Assigned three simple experiments that demonstrate the role of enzymes in digestion. Students examined the breakdown of protein by pepsin, fat by pancreatic lipase, and starch by amylase.	PL
Nervous system	Examined principles of feedback and reflex physiology. Students observed how a thermostat regulates the temperature of a water bath. Students also examine human knee-jerk and pupil dilation reflexes.	PL
Respiration and circulation	Collected students' respiratory volumes, breathing rates, blood pressures, and heart rates under various conditions.	PL
Natural selection	Simulated natural selection by playing the role of a predator selecting various colored paper dots from a colored background. Surviving dots "reproduce" and "predation" was repeated. Students observed the change in frequencies of colored paper dots over several generations.	PL
Taxonomy	Reviewed taxonomic rules of classification to identify several unlabeled animal specimens by phylum and class.	PL
Animal behavior	Required to students to record the behavior of beta fish under various conditions.	PL

Assessments. There were two sources of assessment data for this project: course grades were tabulated for every offering of the course and student surveys were administered at the beginning and end of each quarter. Students took the pre- and post-course surveys through the online course management system and bonus points were offered to students who completed these surveys. The surveys were designed to address: (1) attitudes about the study of biology; (2) knowledge of central concepts (i.e., the process of evolution); and (3) their ability to design and carry out experiments. A copy of the full student survey is available here: <http://tinyurl.com/vlab-eport>.

To measure shifts in attitudes, we employed the Colorado Learning Attitudes about Science Survey for Biology (CLASS-Bio, Semsar, Knight, Birol, & Smith, 2011) designed to measure novice-to-expert-like perceptions about biology. Students rated their agreement on a five-point Likert scale ("strongly agree" to "strongly disagree") to statements such as "Learning biology changes my ideas about how the natural world works." CLASS-Bio is graded by comparing the shifts in student responses to expert consensus (responses from biology PhDs detailed in Semsar, Knight, Birol, & Smith, 2011). There are seven sub-categories of the CLASS-Bio statements revealed through iterative reduced-basis factor analysis (described in Adams et al., 2006): problem-solving difficulty, problem-solving effort, problem-solving strategies, conceptual connections, real world connections, reasoning, and enjoyment.

In order to measure students' conceptual learning, we chose evolution as a domain because it is covered in all three types of laboratory assignments (PL, VL-A, and VL-H). Students read a short vignette (about lizards or canaries) and were asked seven questions about the process of evolution in the given scenario. The vignette and questions were adapted from the Natural Selection Concept Inventory (Anderson, Fisher, & Norman 2002). Students received one vignette for pre-test and a different vignette for the post-test (they were randomly assigned to one of two vignette orders).

To measure students' abilities to carry out research methods, four questions were adapted from the Biological Concepts Instrument (Klymkowsky, Underwood, & Garvin-Doxas 2010). Two questions

were multiple choice and two questions were open-ended questions. Open-ended answers were graded by research assistants who were blind to which type of lab experience the respondent had.

The post-instruction survey also included a few questions regarding how the student accessed the labs and their opinions on various aspects of the laboratory portion of the course. Students in the PL group were queried about group work. Students in the VL-A group were asked to rate their experience with Smart Science Labs (SSL) and Biology Labs On-Line (BLOL). The VL-H group was queried about BLOL and group work. Students provided ratings (on a slider survey item from 0 to 99) of how much they thought a particular component of their lab experience (group work, SSL, or BLOL) contributed to their (1) overall understanding of biology, (2) understanding of science experiments and lab work, and (3) appreciation of biology as an interesting and relevant discipline. Students also provided an overall rating of how much they liked that component by assigning a rating of 1-5 stars.

Results

Final Course Grades

As shown in Figure 1, there were statistically significant differences in the student course grades for the three types of laboratory formats, $F(2, 712) = 55.64, p < .001$. Fisher's LSD corrected post-hoc analyses showed that the students taking the courses with VL-H achieved significantly better grades than PL, $p < .001$, and VL-A courses, $p < .001$. There were no significant differences between PL and VL-A grades, $p = .28$.

We also examined the proportions of repeatable grades (defined as a grade below C or a withdrawal from the course) and the results are depicted in Figure 2. As a complement to the VL-H group's higher GPA, there were significantly fewer repeatable grades in the VL-H course, $\chi^2(2, N = 748) = 27.59, p < .001$. Interestingly, the VL-A group showed the highest proportion of repeatable grades, although the mean course GPA was not different from the PL group.

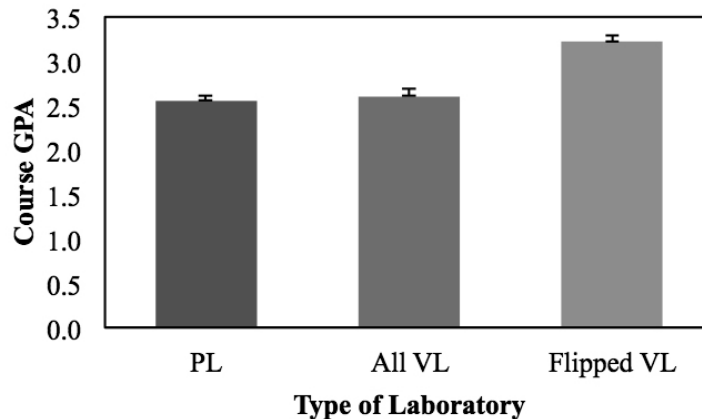


Figure 1. Mean grade point average (and standard error) for each of the three types of lab experience.

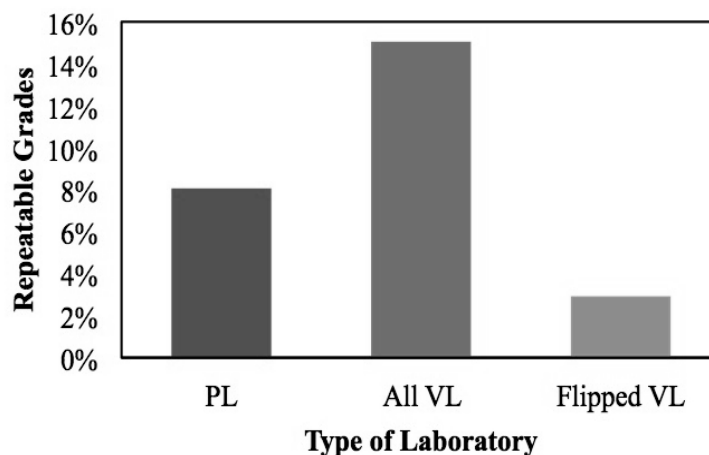


Figure 2. Proportion of repeatable grades for each of the three types of lab experience.

Survey Completion

Because it is difficult to control how seriously students take online assessments, it is important to examine the rates of survey completion across the three types of laboratory formats. In general, pre-instruction survey completion rate (85%) was higher than post-instruction (58%). A chi-square test of homogeneity revealed this ratio did not differ significantly across the three lab types, $\chi^2(2, N = 1075) = 1.72, p = .58$. We then excluded responses where students were unlikely to be reading the prompts (modeled after Semsar, Knight, Birol, & Smith, 2011). Responses were excluded for one of the following reasons: (1) providing the same Likert-scale response (e.g., all “strongly agree”) for more than 90% of statements, (2) incorrectly responding to a statement embedded in the survey (“We use this statement to discard the survey of people who are not reading the questions. Please select ‘agree’ (not ‘strongly agree’) for this question to preserve your answers.”), and (3) for not responding to both pre- and post-instruction surveys. If a student submitted more than one set of acceptable responses, only the first completed response was included in the analysis. There were 343 participants who met all of these criteria and comprised the set of data analyzed for the following survey results.

Student Attitudes toward Biology

The pre-/post- surveys allowed an assessment of the changes in students’ attitudes toward biology for the three types of laboratory formats. For each statement, a student’s shift in response was designated as favorable (agreeing with the expert consensus—not necessarily agreeing with statement), unfavorable, or neutral as detailed in Semsar, Knight, Birol, and Smith (2011). An ANOVA revealed significant differences among the laboratory groups, $F(2, 340) = 4.2, p = .016$. In Fisher’s LSD post-hoc analyses, only the VL-H group showed a statistically significant positive increase in the percentage of favorable responses compared to PL, $p = .008$, and VL-A, $p = .003$. The small negative changes for the PL and VL-A groups were not significantly different from zero, $p = .65$. Table 3 shows the changes in student attitudes towards biology overall as well as broken down by sub-category. These results suggest that the VL-H format has the potential for increasing students’ attitudes towards problem solving and their enjoyment of biology.

Table 3: Mean and Standard Errors for the Change in Percent of Favorable Responses

Category	PL (n=92)	VL-A (n=95)	VL-H (n=156)
Overall	-0.26% ± 1.99%	-1.22% ± 1.55%	+4.57% ± 1.36%*
Problem-solving difficulty	-0.02% ± 2.69%	+0.74% ± 2.97%	+4.28% ± 2.14%*
Problem-solving effort	+0.08% ± 2.38%	+1.32% ± 2.97%	+6.97% ± 2.39%*
Problem-solving strategies	-0.26% ± 2.38%	+4.17% ± 3.59%	+8.39% ± 2.91%*
Conceptual connections	-4.04% ± 2.58%	+1.02% ± 2.74%	+3.21% ± 2.12%
Real world connections	-0.24% ± 2.62%	-2.43% ± 2.82%	+5.25% ± 2.03%*
Reasoning	-6.68% ± 2.80% ^o	-4.41% ± 3.59%	+2.44% ± 2.29%
Enjoyment	+0.97% ± 2.15%	+1.33% ± 2.62%	+6.30% ± 1.95%*

* statistically significant positive change

^o statistically significant negative change

Knowledge of Evolution and Research Methodology

The assessment of content knowledge in the pre/post surveys focused on evolution by natural selection—a topic covered in PL as well as the VL-A and VL-H assignments. None of the three lab formats exhibited significantly different changes in knowledge of evolution, $F(2, 380) = .47, p = .62$. Furthermore, these changes did not significantly differ from zero, p -values $> .4$. We examined performance on the research methodology questions on the pre-post surveys and found a similarly disappointing result. There were no significant differences across the three lab formats, $F(2, 271) = .37, p = .69$, nor any significant differences from zero, p -values $> .2$. It does not appear that the laboratory exercises (nor lecture material) had a long-term positive impact of the students' knowledge of evolution nor research methodology as measured by these questions.

Student Opinions

Because students were queried about specific aspects of the lab section and these components were confounded with lab modality, there were no measures comparing all three groups. Table 4 summarizes student ratings of how much their respective lab sections contributed to their (1) overall understanding of biology, (2) understanding of science experiments and lab work, and (3) appreciation of biology as an interesting and relevant discipline (ratings ranged from 1-99). Their overall rating of that component is summarized as well (ratings ranged from 1-5 stars). We can make a few pairwise comparisons to aid future designs of virtual labs.

The lab section of the VL-A course was comprised of two different kinds of VLs: SSL and BLOL. All of their ratings of SSL were significantly lower than their ratings of BLOL, t -values $> 6.6, p$ -values $< .001$. Students believed that the BLOLs made a more significant contribution to their learning and gave it an overall higher rating.

The VL-H version of the course rated the BLOLs and group work. Student reports of the contribution of these two components to their overall learning of biology, understanding of science, and appreciation of biology were not significantly different from one another, t -values $< 1.6, p$ -values $> .1$. But the VL-H students gave BLOL an overall higher rating than group work, $t(197) = 3.35, p = .001$.

Table 4: *The Means (and Standard Deviations) for Students' Ratings*

Lab Modality	SSL	BLOL	Group Work
PL (n=92)			
Biological understanding			61.96 (23.71)
Scientific understanding			67.07 (24.81)
Appreciation for biology			64.22 (25.36)
Overall rating			3.70 (.97)
VL-A (n=95)			
Biological understanding	44.2 (36.68)	68.33 (24.13)	
Scientific understanding	47.36 (38.3)	69.66 (24.57)	
Appreciation for biology	45.33 (37.29)	66.6 (26.21)	
Overall rating	2.66 (1.66)	4.02 (1.1)	
VL-H (n=156)			
Biological understanding		65.95 (27.19)	64.97 (27.78)
Scientific understanding		67.65 (28.26)	66.1 (28.56)
Appreciation for biology		66.4 (28.62)	64.57 (29.56)
Overall rating		4.03 (1.09)	3.71 (1.24)

Costs Comparison

The final goal was to conduct a cost comparison between lab types. Each PL section accommodates 20-25 students and requires an instructor. At Cal State LA (quarter system), the PL version of this course is typically a large lecture offered with 8 lab sections of 24 students each. Personnel needs are 22 units for instruction (6 units for large lecture and lab coordination, 16 units for labs) and 5 hours of graduate assistance (GA). Using a rate of \$1105/unit instruction and \$14.80/hr./GA, the total cost is \$25,050 to run and teach the course. A course of the same size with VL-A, assuming 30 hours of GA help for drop-in assistance and grading, costs \$11,070. Thus a course taught with VL-A could be offered at 44% of the PL cost with no impact on physical lab facilities. A similar calculation for VL-H would be 71% of the traditional cost and would double the throughput capacity of the physical lab facilities. A comparison of these costs divided by number of students served is shown in Figure 3.

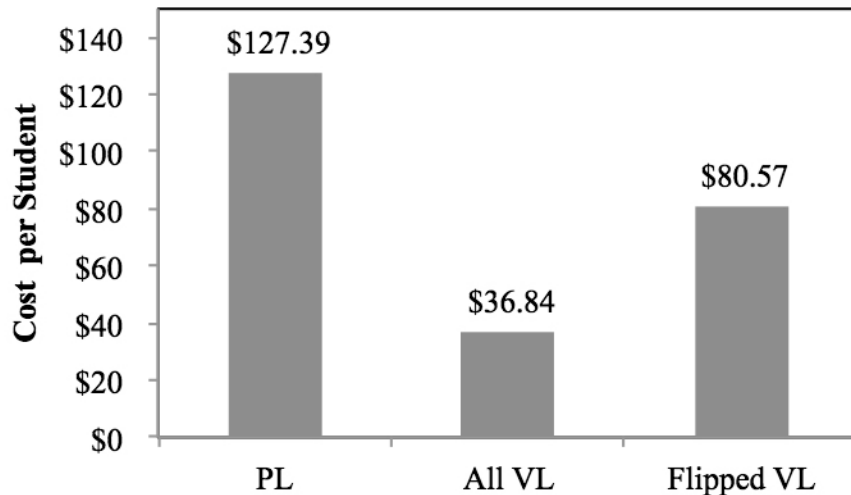


Figure 3. Per student costs for the three lab formats assuming a PL course with 8 labs sections of 24 students per section.

Discussion and Conclusion

This study was designed to explore VLs as a solution to both pedagogical and resource bottlenecks in providing undergraduates with a broad science education. The flipped VL-H model resulted in better grades and more positive attitudes towards biology. Without the alternating in-person labs, the VL-A model resulted in more repeatable grades. In general, the three lab implementations did not foster measurable changes in learning concepts on course-independent concept inventories. The integration of VLs reduced costs and allowed more than twice as many students to enroll compared to the previous PL only year.

These results show some signs of the potential of VLs for improving the GE biology experience when implemented with a longer time devoted to each topic, fewer lab topics, and an equal amount of time in-person. By providing a few, targeted opportunities to actively engage in the conduct of science (formulating hypotheses, designing and carrying out experiments, analyzing and interpreting results) and to meet together to work out any misunderstandings, students seemed to gain a better appreciation of biological science as a process for investigating the natural world. These changes in attitudes may be potentially important for engaging students in further scientific inquiry. The lower grades in the VL-A model could serve as a warning: large-scale implementation without adequate support may lead to worse outcomes for achievement.

The two different implementations of VLs illustrate that the particular labs and how these labs are implemented may make a difference in student attitudes and grades. Student opinions from the VL-A group revealed a preference for one of the types of VLs used in our study (Biology Online Labs). As instructors and departments grapple with the use of VLs, they must grapple with the pedagogical benefits and limitations of different VLs. Our results suggest that these differences in the quality of VLs could be related to student's perceptions of learning and grades. As VL environments continue to be developed, practitioners and decision-makers require data to inform which of these should be adopted in classrooms.

Limitations

A number of caveats limit the generalizability of these data. This was not a randomized controlled trial, the different implementations were taught by different instructors and offered in different quarters and thus the findings cannot be generalized to other contexts. However, when we analyzed the subset of VL students that were taught by the same instructor (VL-A versus one of the VL-H sections), we predominantly obtained the same results as in the larger analyses¹. Also, the PL and VL assignments largely differed in their content and focus because the assignments highlighted different affordances present in the two lab environments. For instance, VL assignments tended to emphasize designing experiments (and often conducting multiple experiments) while PL assignments tended to emphasize

¹ For the following analyses we only focused on the VL-A (N = 181) and VL-H (N=172) taught by the same instructor. Only 184 students completed all measures (94 from the VL-A group and 90 from the VL-H group) and there were no statistical differences between the groups in terms of their rates of completion, $\chi^2(1, N = 353) = .01, p = .94$. There were no significant differences in knowledge of natural selection/research methods, $F(1,182) = 1.4, MSE = .06, p = .24$. In terms of course GPA, there was a significant difference that matched the pattern found in the larger sample where the VL-H group had higher average grade ($M = 3.36, SD = .82$) than the VL-A group ($M = 2.64, SD = 1.04$), $F(1, 344) = 51.46, MSE = .88, p < .001$. Attitudes in this smaller sample shifted in a qualitatively similar manner to the greater sample. In the VL-A group, there were no statistically significant changes in any of the CLASS-BIO subscales. In the VL-H group, there were significant positive changes in attitude overall, as well as in problem solving difficulty, problem solving effort, and enjoyment. The only significant pattern that was not replicated in this smaller sample was the significant positive shift in problem solving strategies ($p = .2$). This subgroup also rated SmartScience Labs as significantly worse (both in overall ratings and contributions to learning) than Biology Labs Online and group work. They did not show any significant differences in opinion between Biology Labs Online and group work. All other patterns found in the main analyses were confirmed in this subgroup that had the same lecture instructor with different lab modes.

measurement and other lab skills that can be easily implemented in a physical environment. In doing so, this study compared lab implementations that would make the most sense in each lab environment.

Part of the benefit of the VL-H model may have been that there were fewer topics covered and more time was devoted to each topic. This more focused approach may have contributed to students' shifts in attitude. Future research may want to address whether focusing on fewer topics in a PL setting would foster similar improvements in attitude and grades.

Disappointingly, students from all three models showed no measurable differences in their competence on the concepts questions. Given that these quizzes were not tied to their grade in any way, students may not have been motivated on the online assessment. Also, because these concept inventories were designed to be only analogous to the situations they encountered in class, the questions were superficially different from questions they may have answered in class. Much of the research in cognitive science suggests that, especially for novice learners, superficial differences (e.g., an evolution question about canaries versus lizards; in class versus online) prevent students from transferring their knowledge from one situation to another (see Barnett & Ceci, 2002 for a review). Students also answered post-instruction questions several weeks after the relevant lab assignment so the delay may have also influenced their ability to demonstrate learning. Although we knew that these superficial differences and the time delay would presumably lower students' transfer of their learning from the course to the quiz, such "far transfer" questions are considered the gold standard in many studies of learning. If the objective of a GE course is to impart flexible use of biology concepts when encountering other real life situations, these superficially different questions would be able to detect an effect on that type of ideal generalizable learning.

Ideally, one might propose a version of this study that would be more controlled in terms of the content of the lab assignments. However, this would be a difficult and potentially invalid test of what departments and professors would normally do when designing a lab course. Typically, when instructors/departments are charged with teaching a lab course, they would look for the best materials to meet their teaching goals that would also meet the demands of resources (e.g., physical space, budgets). If the focus of a course were to improve students' skills with laboratory equipment or measurement, it would be difficult to find a VL equivalent of PL experience. If the teaching goals emphasize conducting experiments that would provide evidence for natural selection or ecological population dynamics, it would be difficult to implement the PL equivalent of a VL experience. However, we hope that studies like ours will motivate other research groups to collect more refined evidence in large scale implementations of PLs and VLs in science courses. Also, as the technology of VLs improves, we may be able to conduct more closely aligned comparisons of VL and PL courses.

Replacing traditional labs with VLs may seem like a pedagogical risk. But "traditional" does not necessarily mean that the model is best for learning and changing student attitudes. Studies like this one are intended to spur instructors on in the process of improving pedagogy rather than conclude that one model is the best. Given that universities and departments struggle with resource limitations, these large-scale comparisons provide useful data about the benefits and limitations of VL implementation. But the results of this implementation show that there is much work still left to do.

References

- Anderson, D.L., Fisher, K.M. & Norman, G. J. (2002). Development and evaluation of the Conceptual Inventory of Natural Selection. *Journal of Research in Science Teaching*, 39, 952–978.
- De Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science*, 340, 305-308.
- De Jong, T. (2006). Computer simulation—Technological advances in inquiry learning. *Science*, 312, 532-533.
- De Jong, T. (2010). Learning by creating and exchanging objects: The SCY experience. *British Journal of Educational Technology*, 41, 909-921.
- Finkelstein, N.D., Adams, W.K., Keller, C.J., Kohl, P.B., Perkins, K.K., Podolefsky, N.S., Reid, S., & LeMaster, R. (2005). When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. *Physical Review Special Topics—Physics Education Research*, 1, 010103.
- Hennessy, S., Deane, R., & Ruthven, K. (2006). Situated expertise in integrating use of multimedia simulation into secondary science teaching. *International Journal of Science Education*, 28, 701–732.
- Jacobson, M.J., & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and research challenges for the learning sciences. *Journal of Learning Sciences*, 15, 11–34.
- Klymkowsky, M W., Underwood, S.M., & Garvin-Doxas, K. (2010). Biological Concepts Instrument (BCI): A diagnostic tool for revealing student thinking. *arXiv* 1012.4501.
- Klahr, D., Triona, L.M., & Williams, C. (2007). Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. *Journal of Research in Science Teaching*, 44, 183–203.
- Linn, M. C., Lee, H. S., Tinker, R., Husic, F., & Chiu, J. L. (2006). Teaching and assessing knowledge integration in science. *Science*, 313, 1049-1050.
- National Research Council. (2006). America’s lab report: Investigations in high school science. Washington, DC: The National Academies Press.
- National Research Council. (2007). Taking science to school: Learning and teaching science in grades K-8. Washington, DC: The National Academies Press.
- Rutledge, M.L. & Sadler, K.C. (2007). Reliability of the Measure of Acceptance of the Theory of Evolution (MATE) instrument with university students. *The American Biology Teacher*, 69, 332–335.
- Semsar, K., Knight, J.K., Birol, G. & Smith, M.K. (2011). The Colorado Learning Attitudes about Science Survey (CLASS) for use in biology. *CBE Life Sciences Education*, 10, 268–278.
- Tobin, K. (1990). Research on science laboratory activities: In pursuit of better questions and answers to improve learning. *School Science and Mathematics*, 90, 403-418.
- Beltz, D., Deshamais, R., Narguizian, P., & Son, J. (2016). Comparing Physical, Virtual, and Hybrid Flipped Labs for General Education Biology. *Online Learning* 20 (3) 228 - 243.

- Triona, L., & Klahr, D. (2003). Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments. *Cognition and Instruction, 21*, 149–173.
- Welch, W. W., Klopfer, L. E., Aikenhead, G. S., & Robinson, J. T. (1981). The role of inquiry in science education: Analysis and recommendations. *Science Education, 65*, 33-50.
- Windschitl, M., & Andre, T. (1998). Using computer simulations to enhance conceptual change: The roles of constructivist instruction and student epistemological beliefs. *Journal of Research in Science Teaching, 35*, 145-160.
- Winn, W., Stahr, F., Sarason, C., Fruland, R., Oppenheimer, P., & Lee, Y-L. (2006). Learning oceanography from a computer simulation compared with direct experience at sea. *Journal of Research in Science Teaching, 43*, 25–42.
- Zacharia, Z. C., Olympiou, G., & Papaevripidou, M. (2008). Effects of experimenting with physical and virtual manipulatives on students' conceptual understanding in heat and temperature. *Journal of Research in Science Teaching, 45*, 1021-1035.
- Zacharia, Z.C. (2007). Comparing and combining real and virtual experimentation: An effort to enhance students' conceptual understanding of electric circuits. *Journal of Computer Assisted Learning, 23*, 120–132.