Improving the Design of Undergraduate Biology Courses toward the Goal of Retention: The Case of Real-World Inquiry and Active Learning through Metagenomics

Amanda M. Cottone* and Susan Yoon
Graduate School of Education, University of Pennsylvania, Philadelphia, PA 19104

Retention in science is low in undergraduate populations, especially for under-represented minority (URM) and first generation (FG) college students. Thus, educators have been called upon to design curricula to counteract this trend. This study examined variables most likely to lead to retention, such as increased achievement, improved attitudes, and self-efficacy beliefs, through participation in active learning and real-world research experiences in an introductory biology course. The research experience was embedded in metagenomics content and processes that have increasingly gained focus in microbiology. This study also investigated differences in learning outcomes when the curriculum was infused with more active learning. The active learning components included integrating interactive technology into the pre-lab lectures, providing students with authentic protocols to conduct lab work, and allowing students to rerun problematic samples. Results showed increased achievement for URM/FG students, although this was not strongly tied to the active learning elements incorporated into the three-week metagenomics research experience. However, students participating in research with more active learning did report higher frequencies of engaging in mastery experiences (an important source of self-efficacy) when compared with students engaged in research with less active learning. This analysis can aid in identifying specific curricular design features associated with promoting retention in undergraduate biology and science programs in general.

INTRODUCTION

Participation in the science workforce in the United States is low (1), and this can partially be attributed to low retention rates of undergraduate students (2). These low retention rates are further exacerbated for underrepresented minority (URM) and first-generation (FG) college students pursuing science, and the existence of achievement gaps within these groups has been well documented (3, 4). These results exist despite significant financial investment—across local and national levels—over the past decade in programs designed to address and mitigate such achievement gaps (5, 6). One likely explanation for this lies in the passive-learning approach found so commonly in large enrollment introductory courses. Undergraduate educators have been called upon to reformulate these courses to encourage active learning (AL), collaboration, cross-disciplinary connections, and higher-level problem-solving skills (7, 8). With the goal of implementing an intervention that would increase student retention in science, we explored the impact of incorporating a metagenomics research curriculum for introductory undergraduate biology students majoring in science and examined the effects of increasing AL in the curriculum (9, 10).

Metagenomics is a fast-growing and diverse field that incorporates cutting-edge sequencing technologies to improve our understanding of microbial genomes, with potential applications spanning various fields, such as medicine, alternative energy, and agriculture (11, 12). We hypothesized that by using current and authentic scientific applications and problem-solving techniques, students would develop an interest in what they were studying that would in turn improve their achievement and self-efficacy. We used a mixed methods approach to investigate the following questions:

1) How does infusing more active learning components into the curriculum benefit the achievement of URM/FG students when compared with their counterparts?
2) To what extent do students perceive the metagenomics research curriculum as authentic and/or incorporating elements of a mastery experience?
Theoretical considerations

Undergraduate research experiences allow students to engage in authentic scientific practices and help increase retention because they promote students’ self-efficacy and their ability to tolerate obstacles related to doing and learning science (13, 14). Students also incur benefits through to graduation and beyond, in their professional careers (15). Course-based research experiences have the added advantage of being available to all students rather than just those who seek extracurricular opportunities (16).

Self-efficacy theory (17) describes how belief in one’s ability to learn is an essential component of the continued pursuit of learning within a subject. Studies in science education have identified science self-efficacy as an important mediator between research experiences and retention (e.g., 13). The main mechanism by which self-efficacy is increased is through mastery experiences, which are based on participation in authentic practice (17). Perceived authenticity is one element of a mastery experience that has been widely researched (9, 16). In addition to authentic practice, other elements include the ability to succeed in obtaining and interpreting results and developing personal beliefs about one’s capability to engage in subsequent activities within a subject (18).

However, it is important to note that the design of research curricula can vary widely across contexts, and there can be elements of “cookbook laboratory” exploration that can still be labeled as research experiences (16). Thus, while the curriculum may be considered authentic research from the perspective of the designer, the student perspective may be quite different, especially when they simply follow prescribed steps in a traditional lab protocol (19, 20).

To address this potential limitation, as well as improve learning outcomes for students, we infused more AL components into the research experience. Research has shown that the incorporation of AL in traditional lecture courses can increase retention, achievement, and engagement (3, 21). Here we investigated whether increasing AL in a research-based lab curriculum promoted similar outcomes.

METHODS

Context

The metagenomics research module spanned three weekly lab sessions over the course of the spring and fall 2018 semesters. Students conducted research and engaged in analytical practices in metagenomics to characterize and test hypotheses about the leaf microbiome of trees in a nearby urban park. The research questions, hypotheses, and study design were formulated beforehand for the students. The general workflow is outlined in Figure 1.

In Session 1, students were given frozen leaf samples previously collected at the study site. They first isolated all DNA from the leaf, which included both bacterial and leaf DNA, and then used PCR to amplify a region of the 16S rRNA gene. In Session 2, students then purified and reamplified the target 16S DNA, which was then sent for next generation sequencing (NGS). Students understood that NGS is a high-throughput sequencing approach that has revolutionized microbiology. In Session 3, students analyzed their metagenomic libraries and compared their findings with libraries generated from the same trees in previous semesters. The goal of the sequence of these sessions was to provide students with a sense of how scientists have conducted and analyzed the latest research on the leaf microbiome with sequencing techniques that have more recently emerged.

Students were not cultivating bacteria so there were no biosafety handling issues. When working with the chemicals in the extraction (Qiagen #13400: https://www.qiagen.com/us/resources/resourcedetail?id=06aeba19-1108-4ce1-b036-6a13153b0f5b&lang=en) and purification (Qiagen #28104: https://www.qiagen.com/us/resources/resourcedetail?id=95f10677-aa29-453d-a222-0e19f01ebe17&lang=en) kits, students wore suitable lab coats, disposable gloves and protective eyewear. The exact procedures and handbook for these kits can be found at the URLs provided.

The AL activities comprised aspects of the 7E instructional model that promote inquiry-based learning practices (22–24). This model engages students in exploratory and cognitively rich learning strategies that are different from didactic instruction. Table 1 shows the features of the 7E model intentionally manipulated in the study’s experimental design. These manipulations include: 1) the use of authentic research protocols, 2) incorporating interactive technology in the pre-lab lectures, and 3) allowing students the opportunity to repeat problematic DNA extractions and gel electrophoresis analyses. We removed the DNA extraction and gel purification protocol from the lab manual and instead instructed students to work directly from the protocols written by the manufacturer. In addition, students using

![FIGURE 1. Workflow of leaf metagenomics curriculum.](image-url)
the authentic protocol were prompted to make their own decisions regarding amounts and types of reagents used. An interactive polling technology was used by the lab instructors to allow students to directly participate in pre-lab lectures using their personal devices. Such platforms can more effectively elicit students’ prior knowledge and formatively assess their knowledge. The third AL design feature for students in the experimental group was the ability for them to troubleshoot unexpected results. During Session 2 of the research curriculum, these students used gel electrophoresis techniques to decide whether the amplification of target bacterial DNA was successful. If not, students were given the option to either set up their initial PCR a second time or repeat the DNA extraction before moving onto the gel purification procedure. This design feature helped to shift decision-making responsibility to the students and allowed them to practice troubleshooting issues that arise—a practice common in research but not in coursework.

**Participants**

All students enrolled in the second semester of their introductory-level biology lab series (247, of whom total participants=157) engaged in a research experience focusing

---

**TABLE 1.**

Quasi-experimental design of this study using the 7E model, where elements of the elicit, engage, and explore categories were manipulated.

| Learning Instructional Model 7Es | Comparison Group (Sp18, n=120 Participants) | Experimental Group (Fa18, n=37 Participants) |
|----------------------------------|---------------------------------------------|---------------------------------------------|
| Elicit and Engage                | Students read detailed lab manual chapter on concepts, terminology and procedural steps related to metagenomics + low stakes pre-lab quiz on pre-lab reading taken outside of class + pre-lab lecture that largely recapitulates the lab manual content. | Students read abridged lab manual chapter (includes link to authentic protocols) + low stakes pre-lab quiz on pre-lab reading taken outside of class + active learning embedded into pre-lab lectures (via the software Pear Deck and Poll Everywhere) to reinforce learning goals and to also present content which would otherwise be included in cookbook lab manual. |
| Explore                          | Students are given necessary supplies to conduct metagenomic research using a step by step written protocol for DNA extraction and gel purification procedures specially written for students. Students run experiments, but there is no opportunity for them to troubleshoot and re-run a procedure if they do not achieve the target PCR product. | Students are given necessary supplies to conduct metagenomic research using authentic protocols for DNA extraction and gel purification procedures. Students run experiments and determine whether they achieved sufficient product on the first attempt; if unsuccessful, they are given the opportunity to troubleshoot and re-run the procedure. |
| Explain                          | During the third lab session, students are given a chart that walks them through the data analyses generated by all lab sections combined (each row refers to a graph or table) and prompts them to determine whether the data support the hypothesis and explain how. Then students justify in 3 to 5 sentences whether they would argue that their research hypotheses were supported and why. | Same assessment as comparison group. |
| Elaborate, Evaluate and Extend    | Student pairs write a short paper that explains the broader implications of metagenomics research for society and human life (using current research papers from an area of their interest) and for the leaf microbiome (using current research papers related to leaf microbiome). Next, pairs are assigned a metagenomic research paper and will check off and explain whether the study addressed important elements in experimental design (e.g., 200 human gut samples; 0 controls; repeated sampling over time; etc.). At the end, students explain whether the study is deficient and how they could improve it. | Same assessments as comparison group. |
on metagenomics. Participants’ class standings are presented in Table 2. All participants were either declared or intended STEM majors. There were seven lab instructors who implemented the research curriculum. Teaching experience varied between instructors and was categorized as High (>5 years experience teaching lab courses), Medium (2 to 3 years experience), or Low (<2 semesters of experience). Lab instructors met with the laboratory course coordinator (~2 h weekly over 3 weeks) for training on implementing the research curriculum and were given the lab chapters and teaching notes to read beforehand. Teaching notes contained suggested workflow, relevant background to cover, and procedural guidance.

| Class Standing | Comparison | Experimental |
|----------------|------------|--------------|
| Freshman       | 64% (92)   | 0% (0)       |
| Sophomore      | 24% (35)   | 57% (24)     |
| Junior         | 10% (14)   | 33% (14)     |
| Senior         | 3% (3)     | 5% (2)       |
| Post-baccalaureate | 0% (0) | 5% (2)       |

**TABLE 2.**
Class standings of participating students between the two treatment groups.

Data sources and analysis

We used R v3.5.3 (25) to run all statistical analyses in this study. We calculated the Gain Score (# correct post – # correct pre) of 12 multiple-choice metagenomics content survey questions to compare content knowledge gains between groups. We also used the Final Score students received in their introductory biology course as another outcome variable. Final scores represent the combined grade of both lecture and lab assessments for the course. It is important to note that all participants within the experimental group had AL embedded into the semester-long lecture component of the course, whereas the comparison group received a didactic lecture format. Therefore, we used this metric to compare potential combined effects of increased AL in both lecture and lab for the experimental group, versus more traditional style learning in lecture and lab for the comparison group.

We ran two separate ANCOVAs to determine whether there was any significant effect of treatment on either of the two outcome variables. In both models we included the lab instructor as a fixed factor to determine whether there were main effects of the instructors on achievement. For the Gain Score model, we also included students’ survey pre-scores as a covariate to control for any differences in prior knowledge. Then we calculated the effect size of the outcome variables to determine the magnitude and directionality of any significant differences from the ANCOVA. Effect sizes were interpreted as small (0.2), medium (0.5), or large (0.8) (26). We collected initial GPAs from students and used ANOVAs to assess whether an achievement gap existed in our study population. Then we ran the ANCOVAs again and included URM and FG status as a factor to investigate how the models changed.

We also conducted post-intervention focus group interviews to elucidate how students perceived the authenticity of the research curriculum, as well as other mastery experience elements and differences across treatment groups. Interviews were recorded, transcribed, and then coded for evidence of mastery experiences that included 1) perceived authenticity of the research (coded as Authenticity); 2) an ability to succeed in obtaining, interpreting and analyzing results (Interpretations); and 3) the development of personal beliefs about one’s capability and willingness to pursue subsequent activities within science (Personal Beliefs). Additionally, while successful mastery experiences predict positive self-efficacy, unsuccessful experiences can generally serve to lower it (18). Therefore, when each of these three elements was articulated, they were coded as positive (Pos) or negative (Neg). There were three more codes that emerged during analysis: Pos Attitudes, Neg Attitudes, and Neutral Personal Beliefs. All eight codes are defined alongside exemplars in Table 3. A participant’s utterance pertaining to either attitude or mastery experience elements represented a coded unit (n=123). Two independent raters were trained and then independently coded 20% of the data (n=25), which returned a Cronbach’s interrater reliability score of 0.97. The remaining codes were coded by the first author.

In order to address limitations in making causal inferences with quasi-experimental designs, we used matching methods to preprocess the dataset and reduce bias (27). Using MatchIt (28), we used a mix of covariates collected from participants in the pre-survey and tested a variety of matching procedures to find the best balance between treatment groups. Nearest neighbor matching, which uses a logistic regression to estimate the propensity score (28), produced the most positive balance improvement (overall bias reduction was 68.4%). The matched dataset resulted in 37 comparison individuals (with 80 unmatched and discarded) and 37 experimental individuals (with 0 unmatched).

**RESULTS**

To address research question 1, we used quantitative analyses to understand the difference in overall achievement gains and how they compared with URM/FG students between the comparison and experimental groups. To address research question 2, we used qualitative analyses to understand the difference in the perception of mastery experience elements between the two groups.

**Overall achievement gains**

Table 4 summarizes the main effects of treatment on Gain Score and Final Score. The first ANCOVA revealed a
TABLE 3.
Codes categorization manual.

| Category and Definition | Example                                                                                                                                                                                                 |
|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| **Pos Authenticity:** Expressed how the metagenomics curriculum felt like a real-life research scenario. | J: I felt like the whole process, as small as it is, of […] downloading the data and getting it up and running and having to go through [it] yourself and being able to […] view it in different ways by […] highlighting the colors that you want […] and by looking at species, or by year. How we did that, I thought that was […] very realistic […] like how it actually is in science plus […] I thought that was interesting. S: I also liked the computer bit, I thought it was really cool and […] it’s definitely really important and should be used more in biology labs, because that’s […] ultimately where the data gets processed. (James and Seeger, Experimental Group, 27 Apr. 2018) |
| **Pos Interpretations:** Technical success (extracting, amplifying, purifying DNA), or articulated successful comprehension in interpreting metagenomic data. | F: Just in general […] being like able to take long papers like that, and […] data, and […] actually analyze them in different ways. For example, we learned about the Simpson’s whatever equation and reading […] the PCoA plots. And just analyzing data in different forms and styles. R: And combining, analyzing different things to form a more cohesive interpretation. (Frankie and Ramona, Experimental Group, 26 Apr. 2018) |
| **Pos Personal Beliefs:** Articulated choice or ambition to continue studying science as affected by the metagenomics research experience. | P: In terms of career though, I’ll say that I’m particularly interested in how this kind of research method might make its way into archaeobotany. Like, when we were domesticating plants, what kinds of microbes were we domesticating along with them? So good questions from my field have come from this experience. (Peter, Experimental Group, 15 Dec. 2018) |
| **Pos Attitude:** Expressed positive opinions or emotions regarding their learning, in a broader sense than what defines a positive mastery experience. | J: In terms of the assessments, I loved the data analysis one. S: mhm. That was great… J: I thought it was also a great idea that that whole week […] in lab was dedicated to [having time] to work on it, to ask questions. S: mhm, cause it was pretty technical J: Yeah. (James and Seeger, Experimental Group, 27 Apr. 2018) |
| **Neg Authenticity:** Expressed negative opinion or emotion regarding a lack of authenticity of the research experience. | A: I think once I do what I am doing this summer; that’s going to feel like my first research experience to me. […] when someone asks “do you have any research experience?” […] Biology lab would not come to mind, I wouldn’t bring up Bio 101 and 102 lab. […] this thing that I’m doing this summer is kind of like the one thing that sticks in my mind that’s a research experience for me. (Angela, Comparison Group, 26 Apr. 2018) |
| **Neg Interpretations:** Expressed lack of confidence in ability to obtain or interpret results regarding the research curriculum. | S: It seems kind of silly […] that freshmen, undergraduate freshmen are going to critique […] the work of some, you know, PhD… C: Critique. Like who am I? S: Yeah exactly. C: [laughing] To critique. S: Yeah, as opposed to maybe to study it, or you know summarize a little bit, or just describe it, learn from it, rather than being like “oh you know this is how your experimental design could have been better.” (Sachel and Charlie, Comparison Group, 26 Apr. 2018) |
| **Neg Attitude:** Expressed negative opinions or emotions regarding their learning, in a broader sense than what defines a negative mastery experience. | C: It was boring. I feel like evaluating data is always boring honestly, which isn’t […] necessarily a bad thing, but […] bringing your conclusion and […] all your information together is not as exciting as actually doing the experiment of course, because you’re reading and your analyzing and that’s […] not as fun. (Charlie, Comparison Group, 26 Apr. 2018) |
| **Neutral Personal Beliefs:** Development of personal beliefs about capability and willingness to pursue subsequent tasks/activities within science occurred from sources/experiences unrelated to metagenomics research curriculum. | A: So I’m doing biology research this summer; I got […] a position at [a research institution] but what made me want to go into research was actually Bio 101 lecture, and […] Professor Watson and his lecture on cancer research and […] the new, today things that are being done and discovered, that’s what made me realize, like wow, science is a now type of thing. It’s not just learning about what Joe Schmo did in 1950 and we’re just learning about the experiment and it means this. It’s things that are continuing to be discovered today, so it was really the lecture more than the lab, I think, that made me interested in research. (Angela, Comparison Group, 26 Apr. 2018) |

Pos = positive; Neg = negative.
non-significant effect of treatment on Gain Score. We still compared means between groups and calculated the effect size. The mean (standard deviation) of the Gain Score for the comparison was 1.83 (2.6) and for the experimental group, 2.92 (2.7), with a small to medium effect size of 0.41. This means that the infusion of more AL perhaps slightly improved students’ content knowledge gains, but this effect was not significant.

The next ANCOVA revealed that treatment did significantly affect Final Score. The average Final Score for the comparison was 87.5 (5.8) and for the experimental group, 91.1 (5.3), with a medium to large effect size of 0.61. This means that the infusion of more AL did significantly improve students’ final score in their biology course. Taken together, these results show that the treatment condition (i.e., more AL) had a slight, yet non-significant effect on students’ metagenomic content knowledge gains and a significant effect on students’ final score in their biology course.

Achievement gap

We tested for the presence of an achievement gap between URM and non-URM students. The average GPA of URM students was 3.39 (0.34), and that of non-URM students was 3.65 (0.37). These means were significantly different \( F(1, 168) = 16.11, p < 0.00001 \). These data suggest an achievement gap between URM and non-URM students exists among science majors at the university.

Next, we tested for the presence of an achievement gap between First-Gen students (neither parents graduated college), One-Parent students (one parent graduated college) and Both-Parent students (both parents graduated college). The mean GPAs of these groups were 3.28 (0.37), 3.34 (0.54), and 3.64 (0.33), respectively, and were significantly different \( F(2, 167) = 10.11, p < 0.00001 \). Post-hoc tests showed that both First-Gen and One-Parent status students had significantly lower GPAs than Both-Parent students, but they were not significantly different from each other. These data suggest an achievement gap existed between students from households with only one or neither parent having graduated from college (hereafter referred to as FG) and students with both parents having graduated college.

Since initial GPA was significantly lower for students of URM and/or FG status, we combined this into one factor (hereafter URM/FG). This decision is based on previous studies that have grouped URM and FG status into a single variable when evaluating achievement gaps, and their remediation through AL strategies in undergraduate science education (3, 21), and the results of this study would therefore be more comparable with existing research studies on this topic.

Achievement gains including URM/FG as factor

Table 5 summarizes the main effects of treatment and URM/FG on Gain Score and Final Score. Neither treatment nor URM/FG had significant effects on Gain Score, but we still calculated effect size for URM/FG students. The mean Gain Score for URM/FG students within the comparison group was 1.81 (2.7) and for the experimental group 3.28 (2.8), with a medium effect size of 0.52.

Treatment and URM/FG did predict a significant main effect on Final Score. The mean Final Score for URM/FG students within the comparison group was 85.3 (6.7) and for the experimental group 90.4 (6.5), with a medium to large effect size of 0.66. This means that the greater infusion of AL in the experimental group significantly improved URM/FG students’ final score in their biology course.

These results show that increased AL seemed to improve but did not have a significant effect on students’ metagenomic content knowledge gains, and this slight effect was larger for URM/FG students (0.51) when compared with the group as a whole (0.41). Moreover, AL did have a strong significant effect on URM/FG students’ final score.

### Table 4

F and p values for factors included in two ANCOVAs show treatment significantly improved students’ final course scores but not metagenomics content knowledge gain scores.

|     | Treatment | Lab Instructor | Pre-score |
|-----|-----------|----------------|-----------|
| Gain Score ANCOVA | \( F(1, 64) = 3.04, p = 0.086 \) | \( F(1, 64) = 0.808, p = 0.525 \) | \( F(1, 64) = 13.47, p = 0.0005 \) |
| Final Score ANCOVA | \( F(1, 66) = 7.61, p = 0.008^* \) | \( F(1, 66) = 1.16, p = 0.340 \) | — |

* indicates a significant effect of treatment on the outcome variable.

### Table 5

F and p values for factors included in two ANCOVAs show treatment significantly improved students’ final course scores and this effect was greater for URM/FG students.

|     | Treatment | URM/FG | Lab Instructor | Pre-score |
|-----|-----------|--------|----------------|-----------|
| Gain Score ANCOVA | \( F(1, 63) = 3.0, p = 0.088 \) | \( F(1, 63) = 0.122, p = 0.728 \) | \( F(4, 63) = 0.784, p = 0.540 \) | \( F(1, 63) = 13.28, p = 0.00054 \) |
| Final Score ANCOVA | \( F(1, 63) = 7.905, p = 0.007^* \) | \( F(1, 63) = 4.552, p = 0.037^* \) | \( F(5, 64) = 1.007, p = 0.418 \) | — |

* indicates a significant effect of treatment or URM/FG status on the outcome variable.
in their biology course, and again, this effect was larger for URM/FG students (0.66) when compared with the group as a whole (0.61).

Qualitative analysis

There were eight focus group participants and they all expressed a commitment to pursuing biology studies and/or more research experience. Subject matter interest varied among participants, with majors in biology, math, medical anthropology, and neuroscience. All participants cited their motivation to pursue science studies as unrelated to the metagenomics research experience (i.e., Neutral Personal Beliefs code).

Comparison group participants (n = 3) averaged 13 comments per student pertaining to attitudes or mastery experiences whereas experimental group participants (n = 5) averaged 16.8 comments. The percent frequency of each code category scored is shown in Figure 2. Both groups reported relatively low frequencies of Negative Attitude reflections (13% for the comparison and 7% for the experimental) yet higher frequencies of Positive Attitude reflections (23% comparison, 36% experimental). Moreover, Pos Authenticity perceptions in the research curriculum seemed to vary between treatments (0% comparison, 13% experimental), as did Neg Successful Interpretations (18% comparison, 8% experimental). These data demonstrate that comparison group participants did not articulate they perceived the research curriculum as authentic, yet all those interviewed from the experimental group perceived some aspect of authenticity. There were also fewer articulations in the experimental group of lack of confidence in the ability to obtain/interpret results regarding the research, yet two of three participants in the comparison group noted this (most often regarding bioinformatics and critiquing primary literature).

A final noteworthy trend emerged with students enrolled in the experimental group. When asked specifically about the use of polling technology incorporated into the pre-lab lectures, four of the five experimental group participants expressed positive attitudes regarding the implementation of this feature in their learning. Students stated that it was helpful to incorporate in that, for example, “[it is very] helpful […] because I think I’ve definitely gone into some labs, […] even having read the lab manual, and not been super sure of what we were doing, so I think it’s nice to have those […] checkpoints.” They also mentioned that seeing other students’ responses and the questions they were also struggling with was helpful for their learning. Moreover, students stated the interactive process helped to keep them paying attention in class, and they would prefer to attend science lectures that incorporated this.

DISCUSSION

These results showed that increasing AL in the research curriculum for URM/FG students had a slight but non-significant effect on students’ metagenomic content knowledge gains; however, increasing AL in lecture/lab combined did have a significant effect on URM/FG students’ final score in their biology course. Active learning during lecture consisted of low-stakes pre-unit quizzes based on the readings, minimal lecturing, and in-class group worksheets and virtual discussion threads on understanding, application, and analysis of concepts. The effect size of these features (0.66) put into context here would translate into a 4.4% increase (or half a letter grade or more) in final score. This is comparable with a previous meta-analysis of various undergraduate STEM courses, where researchers found that the 0.47 effect size of AL over traditional lecturing translated into a 6% increase in exam scores (3).
These findings are important where achievement gaps exist because increased achievement in introductory science courses leads to lower course failure rates and therefore higher rates of retention (3, 29, 30). In this study, URM/FG students had GPAs that were significantly lower (on average by 0.3 points) than their peers. Achievement gaps for URM/FG students have been reported to range from 0.2 to 0.6 grade points (21, 31, 32), and this study population falls within that range. In order to reduce these gaps, curricular interventions need to benefit disadvantaged students to a greater degree (21), and it appears that AL embedded into the lecture portion of the course achieved that, and perhaps so did the AL embedded into the research curriculum, to a lesser degree (i.e., effect sizes for both metrics were greater for URM/FG students when compared with analyses including all students).

The fact that the treatment effect on the metagenomics Gain Score was only slight could be due to the short length of the intervention. A review found that students with research experiences of longer duration tended to demonstrate more positive results in a range of outcomes than students in research programs of shorter duration (29). It is possible that three lab sessions participating in the research curriculum was too short to produce a significant change in content knowledge gains. It is also possible there was a lack of efficacy of the particular AL modifications infused. This finding can point educators to evaluating the outcomes of other AL strategies that are potentially more efficacious over shorter periods. For example, studies have shown that peer discussion and utility value interventions are important measures in promoting achievement in undergraduate science courses (33, 34).

However, qualitative findings showed that all experimental group students perceived some of the research curriculum as authentic, and reflections of this nature were absent for comparison group students. Most students also expressed success and enjoyment in their ability to grapple with the complex nature of bioinformatics (although those in the comparison group did so to a lesser degree). If these two mastery experience elements were perceived differently between the two groups, then students’ self-efficacy is likely to be boosted to a greater degree as a result of the AL elements infused into the research curriculum. The literature on mastery experiences as sources of self-efficacy support this, yet a larger sample would be required to identify whether this trend persists.

On the other hand, the development of personal beliefs was infrequently discussed in interviews, most students describing their beliefs of continuing to pursue science as unrelated to the research curriculum. This suggests that students intended to persist in science regardless of having participated in the research experience. This is consistent with a study on undergraduates working in research programs, where 57% reported that they were committed to graduate school prior to joining a research lab and the position had no effect on their decision (29). In this study, it seemed that by the time students reached their second semester introductory biology course, they had already solidified their commitment to pursuing science. Indeed, there was a high proportion of non-freshmen enrolled in the courses, so perhaps the second semester of the introductory series was too late to address the situation of students who may have been conflicted about persisting.

Other limitations of this study include the evaluation of self-reported outcomes and participation selection bias (9). Only students who elected to take part in the study were included in the analysis, and it is possible that non-participants experienced the course differently than their peers. Another limitation involves the instructor effect and the variability with which the curriculum was implemented. The teaching experience levels suggest there was a range of variability in how lab instructors engaged students in the conceptual material and provided guidance in the execution of the workflow itself. One focus group participant stated, “I think your [lab instructor] can make or break the lab. If you have [an instructor] that doesn’t engage with you, that doesn’t explain anything…it can be harder.” Therefore, instructor variation likely increased statistical noise and made it more difficult to detect and assess true treatment effects.

The improvement in final score for students enrolled in an AL biology course here is likely one essential component in ensuring their future persistence. Further investigation would provide more definitive evidence of persistence (i.e., beyond measures of intended commitment) from longitudinal studies tracking students’ progress and performance beyond undergraduate degree completion, further advancing our understanding of which learning environment features are most critical to increasing retention of URM/FG students in STEM (9, 13). In this study, we present some evidence that the integration of real-world research experiences as well as the deliberate and sustained infusion of AL in lab and lecture courses alike, are potentially important for this goal.

SUPPLEMENTAL MATERIALS

Appendix 1: Metagenomics content knowledge survey assessment
Appendix 2: Focus group interview protocol

ACKNOWLEDGMENTS

This study was supported by the University of Pennsylvania’s School of Arts and Sciences. All data presented for this study have been collected under an approved IRB protocol (#823189), and this research has complied with all relevant federal guidelines and institutional policies. We would like to thank Linda Robinson, Karen Hogan, and Wendy Chan for their support in this in study. The author has no conflict of interest to declare.
REFERENCES

1. U.S. Bureau of Labor Statistics. 2017. Employment projections 2016–2026. U.S. BLS, Washington, DC.
2. President’s Council of Advisors on Science and Technology (PCAST). 2012. Engage to excel: producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Executive Office of the President, Washington, DC.
3. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP. 2014. Active learning increases student performance in science, engineering, and mathematics. Proc Natl Acad Sci 111(23):8410–8415.
4. Harackiewicz JM, Canning EA, Tibbetts Y, Priniski SJ, Hyde JS. 2016. Closing achievement gaps with a utility-value intervention: disentangling race and social class. J Personality Soc Psychol 111(5):745–765.
5. Eagan K, Hurtado S, Chang M, Garcia G, Herrera F, Garibay A. 2015. The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. Front Genet 6:e348.
6. Cook L, Mason M, Morse R, Neuhauser A. 17 May 2016. The 2015 U.S. News/Raytheon STEM Index. U.S. News & World Report, Washington DC. https://www.usnews.com/news/articles/2016-05-17/the-new-stem-index-2016
7. National Research Council. 2009. A new biology for the 21st century: ensuring the United States leads the coming biology revolution. The National Academies Press, Washington, DC.
8. American Association for the Advancement of Science. 2011. Vision and change in undergraduate biology education: a call to action: a summary of recommendations made at a national conference organized by the American Association for the Advancement of Science, July 15–17, 2009. Washington, DC.
9. Gentile J, Brenker N, Stephens A (ed). 2017. Undergraduate research experiences for STEM students: successes, challenges, and opportunities. National Academies Press, Washington, DC.
10. Wieman C. 2017. Improving how universities teach science: lessons from the science education initiative. Harvard University Press, Cambridge, MA.
11. Jurkowski A, Reid H, Labov J. 2007. Metagenomics: a call for bringing a new science into the classroom (while it’s still new). CBE Life Sci Educ 6(4):260–265.
12. Escobar-Zepeda A, Vera-Ponce de León A, Sanchez-Flores A. 2015. The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. Front Genet 6:e348.
13. Chemers M, Zurbriggen E, Syed M, Goza B, Bearman S. 2011. The role of efficacy and identity in science career commitment among underrepresented minority students. J Soc Iss 67:469–491.
14. Hernandez P, Schultz P, Estrada M, Woodcock A, Chance R. 2013. Sustaining optimal motivation: a longitudinal analysis of interventions to broaden participation of underrepresented students in STEM. J Educ Psychol 105:85–107.
15. Rodenbusch SE, Hernandez PR, Simmons SL, Dolan EL. 2016. Early engagement in course-based research increases graduation rates and completion of science, engineering, and mathematics degrees. CBE Life Sci Ed 15(2):ar20.
16. Auchincloss L, Laursen S, Branchaw J, Eagan K, Graham M, Hanauer, Lawrie G, McLinn CM, Pelaez N, Rowland S, Towns M, Trautmann NM, Varma-Nelson P, Weston TJ, Dolan E. 2014. Assessment of course-based undergraduate research experiences: a meeting report. CBE Life Sci Ed 13:29–40.
17. Bandura A. 1994. Self-efficacy, p 71-81. In Ramachaudran VS (ed), Encyclopedia of human behavior, vol 4. Academic Press, New York, NY.
18. Britner S, Pajares F. 2006. Sources of science self-efficacy beliefs of middle school students. J Res Sci Teach 43:485–499.
19. Corwin L, Runyon C, Robinson A, Dolan E. 2015. The laboratory course assessment survey: a tool to measure three dimensions of research-course design. CBE Life Sci Educ 14(4):ar37.
20. Rahm J, Miller HC, Hartley L, Moore JC. 2003. The value of an emergent notion of authenticity: examples from two student/teacher–scientist partnership programs. J Res Sci Teach 40:737–756.
21. Haak D, Hille Ris Lambers J, Pitre E, Freeman S. 2011. Increased structure and active learning reduce the achievement gap in introductory biology. Science 332:1213–1215.
22. Allen D, Tanner K. 2005. Infusing active learning into the large-enrollment biology class: seven strategies, from the simple to complex. CBE Life Sci Educ 4:262–268.
23. Bybee RW. 1997. Achieving scientific literacy. Heinemann, Portsmouth, NH.
24. Eisenkraft, A. 2003. Expanding the SE model. Sci Teach 70:56–59.
25. R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
26. Cohen J. 1977. Statistical power analysis for the behavioral sciences (rev ed). Lawrence Erlbaum Associates Inc, Hillsdale, NJ.
27. Stuart E, Rubin D. 2008. Best practices in quasi-experimental designs: matching methods for causal inference, p 155–176. In Osborne J (ed), Best practices in quantitative methods. SAGE Publications Ltd, Thousand Oaks, CA.
28. Ho D, Imai K, King G, Stuart E. 2011. Nonparametric preprocessing for parametric causal inference. J Statistical Software 42:1–28.
29. Sadler TD, McKinney L. 2010. Scientific research for undergraduate students: a review of the literature. J Coll Sci Teach 39(5):68–74.
30. Springer L, Stanne ME, Donovan SS. 1999. Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology. Rev Educ Res 69(1):21–51.
31. Massey D, Charles C, Lundy G, Fisher M. 2003. The source of the river: the origins, aspirations, and value of freshman at America’s elite colleges and universities. Princeton University Press, Princeton, NJ.
32. Torres KC, Charles CZ. 2004. Metastereotypes and the Black–White divide: a qualitative view of race on an elite college campus. Du Bois Review, Soc Sci Res Race 1(1):115–149.
33. Watkins J, Mazur E. 2013. Retaining students in science,
technology, engineering and mathematics (STEM) majors. J Coll Sci Teach 42:36–41.
34. Tibbetts Y, Harackiewicz JM, Canning EA, Boston JS, Prini-ski SJ, Hyde, JS. 2016. Affirming independence: exploring mechanisms underlying a values affirmation intervention for first-generation students. J Pers Soc Psychol 110(5):635–659.