Course-based undergraduate research is known to improve science, technology, engineering, and mathematics student achievement. We tested “The Small World Initiative, a Citizen-Science Project to Crowdsource Novel Antibiotic Discovery” to see if it also improved student performance and the critical thinking of non-science majors in Introductory Biology at Florida Atlantic University (a large, public, minority-dominant institution) in academic year 2014–15. California Critical Thinking Skills Test pre- and posttests were offered to both Small World Initiative (SWI) and control lab students for formative amounts of extra credit. SWI lab students earned significantly higher lecture grades than control lab students, had significantly fewer lecture grades of D+ or lower, and had significantly higher critical thinking posttest total scores than control students. Lastly, more SWI students were engaged while taking critical thinking tests. These results support the hypothesis that utilizing independent course-based undergraduate science research improves student achievement even in nonscience students.

INTRODUCTION

Course-based undergraduate science research experiences (CURE) include five basic tenets: use of science practices, discovery, broader relevance, iteration, and collaboration (4). These courses have the advantage of reaching many students and have been shown to improve student achievement, critical thinking skills, engagement, retention, and career choices. This is especially true for minority students and first-generation college students (4, 19). Recent studies have shown that the single factor most closely associated with undergraduate science, technology, engineering, and mathematics (STEM) persistence and success for underrepresented students is doing hands-on undergraduate research (1, 4, 14, 19, 26, 27). Other studies, focused on classroom-based research at the introductory level, show evidence that first-year students doing research learn the process of science as well as how scientists practice science (16, 33).

The Small World Initiative for Antibiotic Discovery (SWI) began as the Yale University undergraduate research course Microbes to Molecules in 2013 (29), which was renamed when it was expanded. It is an innovative, classroom-based research experience that focuses on addressing a worldwide health threat: the diminishing supply of effective antibiotics. Students are challenged to discover new antibiotics produced by soil bacteria that they isolate from local habitats (Small World Initiative, www.smallworldinitiative.org). Financial support to expand and disseminate the SWI was provided by the Howard Hughes Medical Institute and the Helmsley Foundation (15, 29). The SWI was first used at multiple schools as a course-based research experience in spring 2014; results reported in the American Society for Microbiology 2014 President’s Address showed many SWI participants scored higher on exams, received better grades, and had higher retention rates than contemporaneous control students (15). When used with nonmajor students, the SWI functions as de facto citizen science, which is defined as having layperson nonscientists do genuine scientific research (5).

Florida Atlantic University (FAU) is a minority-dominant, four-year, research-intensive university (19% Black, 24% Hispanic, and 46% white). FAU has over 21,000 undergraduates with an average ACT score of 22; 77% receive financial aid and 23% receive Pell grants. Over 50% are first-generation college students, and only 41% of first-time-in-college students graduate in six years (www.fau.edu/iea/). FAU is a fertile ground for testing the effect of classroom-based undergraduate research on minority and first-generation college student success.
In May 2014, Dr. Caruso was chosen to go to Yale for hands-on training as an SWI partner instructor. This exposure to several research-intensive undergraduate courses, including the Small World Initiative (30), the Phage Hunters (8), and Genome Sequencing (13), and their promising results, suggested that using research-based inquiry might improve the undergraduate performance of FAU nonscience majors. To assess effects on FAU student performance, we adapted the SWI to the nonmajors Life Science Lab, BSC 1005L. Our hypothesis was that SWI students would earn better lecture grades and have higher posttest critical thinking scores than control students in “business as usual” labs. Most studies of undergraduate science research focus on science student success (4); we have now broadened the studies to analyze effects on nonscience students.

METHODS

IRB and courses

The FAU Institutional Review Board (IRB) approved this study before fall 2014 classes began (FAU IRB #633450-2). Life Science (BSC 1005) is a two-credit lecture course intended explicitly for nonscience majors. Laboratories (BSC 1005L) for BSC 1005 are optional, meet for 110 minutes once weekly, are worth one credit, and have 24 or fewer students per section. In fall 2014, two instructors (Lovelace and Saunders) taught three different lecture sections with >750 students, and in spring 2015, two Instructors (Lovelace and Rowland) taught three sections of similar numbers. Israel taught all SWI and control labs.

Design

Strict randomization of groups was impossible, so this study instead used a quasi-experimental design (7). To minimize any instructor effects, the same teaching assistant taught all experimental and control labs. In both semesters, the SWI lab had the experimental group (n = 22 in fall, n = 21 in spring) and the regular lab section taught immediately before SWI labs was the control group (n = 24 in fall, n = 23 in spring). Both SWI and control labs had near-identical total point grading metrics, with multiple quizzes and homework assignments.

Labs

The control students used the regular Life Science Laboratory BSC 1005L sequence of experimentally-based exercises developed by the FAU laboratory coordinator, which included microscopy, anatomy, fertilization, genetics, physiology, microbiology, neuroscience, conservation, and bioassay exercises (21). Although the labs are based on experiments, there is no independent or inquiry-based research included. Students follow a “recipe” to a known outcome.

Experimental (SWI) lab students were challenged to discover new antibiotics produced by soil bacteria isolated from local environments. Students collected soil samples, isolated diverse bacteria, tested their bacteria against clinically-relevant microorganisms (BSL-1), and characterized those showing inhibitory activity. SWI lab protocols (30) were tested by individual, supervised, undergraduate research students before use in the nonmajors course. Each SWI student had to isolate their own antibiotic-producing bacteria from soil, record and document their results, and submit posters of their findings (in lieu of a final exam).

SWI students were taught lab safety (21) (no violations reported) and how to keep notebooks and test hypotheses. All practices adhered to the ASM Guidelines for Biosafety in Teaching Laboratories (10). These labs do not need a specialized microbiology lab with Biosafety rating because only BSL-1 pathogens are isolated. Specifically, we covered lab equipment and its use, proper behavior, no food or drink, response to emergency situations, and first-aid kit location. We also covered specific topics pertaining to safety in a microbiological lab setting, from proper attire and disposal of materials to aseptic technique and sterilization. We strictly enforced working with lab coats, gloves, long pants, closed shoes, and long hair tied back at all times. Lab coats as well as lab notebooks (and pens) were kept in the lab in order to minimize bacterial exposure outside the classroom (each student was provided with a personal storage container to accommodate their belongings in between lab periods). Additionally, when findings were documented via photographs, smart phones were placed in a ziplock bag.

In the lab, two biohazardous waste containers were present at all times: one large waste bin and one plastic, puncture-proof container specifically for sharps. Any equipment (i.e., erroneously-inoculated plates, toothpicks, bacterial swabs, pipette tips, Eppendorf tubes, etc.) that became contaminated due to contact with bacterial culture (or soil) was disposed of in biohazardous waste. Non-disposable items were autoclaved and sterilized for next usage.

After the termination of each lab period, every lab bench was cleaned with disinfectant to minimize the spread of contaminants (this was also done at the beginning of each lab period). All plates containing samples were collected and placed in their respective ziplock bags (each bench had a gallon-sized ziplock bag in which to store their plates). The instructor stored the plates in the microbiology refrigerator until the next lab period. Any plates necessitating incubation would be incubated for 24 hours and then placed in their respective ziplock bags in the fridge. Transfer of any plate, material, or equipment outside the classroom was performed solely by the instructor. Students only had access to their sample and other materials at their work bench. Students wore protective lab coats and gloves when they handled their bacterial samples at all times. A ventilated hood was used when vortexing a sample was necessary. During the fall (2014) semester, students in groups of six to eight vortexed the sample (contained in Eppendorf tubes) under a hood in the prep room. During the spring semester, the instructors did the vortexing to save time.
SWI students were told there was no way to predict which organisms they would find in soil or how many students would find antibiotic-producing bacteria. They were given sterile 50-mL centrifuge tubes to collect soil. As per the manual, SWI students chose one media and one ESKAPE pathogen (BSL-I) relative (6) against which to test antibiotic-producers for the entire term. Soil samples were diluted in sterile water, vortexed in a Biosafety Level-2 biological safety cabinet, plated on media and incubated at 30°C as described (30), and plates were stored in zipped 4-L plastic bags at 4°C until further use. Students viewed the plates through the bags and confirmed that bacteria were, or were not, producing a chemical (i.e., an antibiotic) that was inhibiting (killing) the tester strain (i.e., an ESKAPE relative) in the bioassay. Photographs were taken with smartphones in zipped 1-L plastic bags and posted on Blackboard and the SWI Facebook page. All gels used SYBR-SAFE rather than ethidium bromide, for safety reasons. This is the last laboratory step the nonmajor students performed. At the end of the semester, the samples were transferred to a research microbiology lab for further testing against a yeast strain, isolation, antibiotic purification, sequencing, or disposal.

Analyses

All SWI and control lab students gave written, informed consent the first day of lab after a permission script was read aloud. The IRB protocol allowed us to offer students up to 1.5% extra credit to take SWI surveys and California Critical Thinking Skills Tests (CCTSTs) (17); lesser amounts were given if students did only pre- or postsurveys and/or CCTSTs.

Online CCTSTs were given as pre- and posttests at the start and end of each semester and pre- and posttest total scores were analyzed by \( \chi^2 \) (2), Kruskal-Wallis non-parametric analysis of variance (KW ANOVA, 7), unpaired t-test (3, 9) and analysis of covariance (ANCOVA, 20), with pretest scores as covariates. \( \chi^2 \) predicts group membership, whether the students will fall into one category or another. The pretest was used to predict high or low class scores as a result of the laboratory experiments. The purpose of the KW ANOVA is to predict variability between two or more groups. We were looking at variability based on examination of the independent variable (the laboratory experiments). We compared both pre- and posttest scores and overall lecture grades using KW ANOVAs. An unpaired t-test compares only two groups and tries to assess the difference between the two groups. An ANCOVA predicts variability similar to an ANOVA, but holds variables constant (in this case, the pretest scores).

Experimental and control group lab grades, lecture grades, and persistence were compared using two-tailed Fisher’s exact test (23), \( \chi^2 \) (2), unpaired t-test (3, 9), KW ANOVA (7), and/or the binomial test (30). Two-tailed Fisher’s exact test reveals the probability of belonging to one group or the other, where there are only two groups to compare. The SWI students were either in the category \( \geq \) C or \( \leq \) D+ during the analysis, compared with control students. Binomial tests are used when there are two possible outcomes. We assessed success or failure via this test and were able to measure the overall probability of success using this result. A binomial test was used to compare lecture mean grades as a result of the SWI initiative.

The first lecture performance measure compared the percentage of students who earned lecture grades \( \geq \) C and \( \leq \) D+ in SWI and control groups. The second lecture performance metric compared mean (X) grades of the SWI and control groups, calculated by assigning points to each grade in the group (\( A = 4, A- = 3.7, B+ = 3.3, B = 3, B- = 2.7, C+ = 2.3, C = 2, C- = 1.7, D+ = 1.3, D = 1, D- = 0.7 \) and \( F = 0 \)), summing the total points of each group and dividing by the \( n \) of students in each lab group. The third lecture performance assessment compared the proportion of SWI and control students whose final lecture grade (%) exceeded or equaled (± 0.6%) the mean lecture grade (X%) in their specific lecture class. Student engagement was measured by tracking length of time spent on CCTSTs, with < 15 minutes spent on the test showing disengagement (17).

Students who got an incomplete grade (I), withdrew, or stopped coming to lecture and/or lab were not included in grade analyses for classes they dropped, but were grouped together as withdrawals for persistence analyses. One spring control student was a Biology major so this student’s data were excluded from all grade analyses. One fall control student and three spring SWI students took lab but did not take lecture, so we were unable to include their lecture grades in analyses. Two spring 2015 control lab students were Honors students who have much higher ACT scores (≥ 28) than the typical FAU student (average 22) and were enrolled in the Honors College (a small residential campus). Their lab and lecture grades were excluded from all analyses since there were no Honors students in the experimental group. Two fall control students stopped attending lab. Two students withdrew from spring SWI lab, one of whom also withdrew from lecture. The spring control lab had one student withdraw from both lab and lecture and one who stopped attending lab and lecture.

RESULTS

Twenty-one of 22 fall SWI students got ≥ C- in lab and all 22 control students got ≥ C in lab, so lab grades were not significantly different. All fall SWI students got ≥ C in lecture, but just 14 fall control students did. Six fall control students earned D+ or below, three failed and one received an I, which became a D- at the end of the next term. When student numbers earning ≥ C or ≤ D+ were compared in separate analyses using Fisher’s exact test (23), fall SWI students had significantly more lecture grades ≥ C and significantly fewer grades ≤ D+ than controls (\( p < 0.0001 \), Fig. 1). When spring SWI lecture and lab grades were compared with spring controls for numbers of students earning ≥ C and ≤ D+ using Fisher’s exact test, as above, they were not
significantly different. When fall 2014 and spring 2015 SWI and control lecture grades were combined, there were 37 students in each group: one SWI student who completed lab failed lecture and another got a D. In contrast, six of 37 control students earned D+ or lower and three students failed. Combined groups tested by Fisher’s exact test (23) for percentages of grades ≥ C showed that significantly more SWI students than controls earned ≥ C (p = 0.026, Fig. 2). A separate KW ANOVA analysis (7) showed that significantly more SWI students earned ≥ C in lecture than did controls (p = 0.0005). Lastly, when combined groups were tested by the binomial test (28), using control results as expected values, SWI students had significantly fewer grades ≤ D+, with a p value approaching 0.

Mean lecture grades in fall SWI and control groups were 2.78 and 2.14, respectively. Fall SWI mean lecture grades were significantly higher than fall controls (p = 0.042) by the binomial test, using the control mean as the expected result (Fig. 3). Mean lecture grades for spring SWI and controls were not significantly different (1.97 and 1.95, respectively, Fig. 3). When fall and spring lecture mean grades were combined and SWI was compared with controls by the binomial test, p approached, but did not reach, significance (p = 0.083).

In fall 2014, seven of 12 SWI students earned lecture grades exceeding or equaling their final section mean percentage grade, while just five of 14 controls did. Significantly more fall SWI students than controls earned a lecture percentage greater than or equal to their final section mean percentage when compared by Fisher’s exact test (p < 0.0001, Fig. 4) (19). In spring 2015, seven of 15 SWI students (i.e., those spring SWI students who took lecture and finished lab) earned a percentage greater than or equal to their lecture section mean percentage while seven of 21 controls who finished lab earned a lecture percentage greater than or equal to their lecture section mean percentage. Differences between spring SWI and control groups approached, but did not reach, significance (p = 0.063 to p = 0.11, several analyses, including χ²).

When fall and spring SWI and control results were combined and proportions of students exceeding or equaling their final lecture mean percentage were compared with the binomial test, using control results as expected values, the proportion of SWI students whose final class percentage exceeded or equaled their lecture section percentage was significantly higher than for controls (p = 0.019, Fig. 4).

In fall 2014, 15 of 22 SWI students (68.2%) and 13 of 24 control students (54.2%) took the CCTST pretest while 19 of 22 SWI students (86.4%) and 16 of the 22 remaining controls (72.7%) took the posttest. In spring 2015, 21 of 21 SWI students (100%) and 22 of 23 control students (95.6%) took the pretest while 15 of 18 remaining SWI students (83.3%) and 13 of 21 remaining control students (61.9%) took the posttest. SWI lab pretest critical thinking total mean scores were not significantly different when both terms were compared (fall 2014 X = 72.2; spring 2015 X = 72.9), and SWI posttest total mean scores increased in both groups to identical levels (X = 73.6). In contrast, fall 2014 control pre- and posttest total mean scores were
identical (68.9), whereas the spring 2015 control posttest total mean score was much lower than its pretest total X (67.2 vs. 71.8). Fall SWI posttest critical thinking total mean scores were significantly higher than controls (SWI X = 73.6 vs. control X = 68.9; \( \chi^2 \) (1), \( p = 0.011 \); unpaired t-test (3), \( p = 0.017 \), Fig. 5).

Fall SWI CCTST pretest mean total scores were higher than control pretest mean total scores and approached significance (0.09 < \( p < 0.10 \), KW ANOVA), so we did an ANCOVA analysis to control for higher SWI pretest mean scores by using them as a covariate (20). This validated the significance of fall SWI students’ higher posttest mean scores as actual learning gains compared with controls (\( p = 0.0069 \), Fig. 5). Spring SWI and control groups’ CCTST pretest scores were similar enough so an ANCOVA was not needed (SWI X = 72.9 vs. control X = 71.8), but posttest total mean scores of spring SWI students were significantly higher than those of controls by unpaired t-test (3) (\( p = 0.016 \), Fig. 5). Combining fall and spring posttest total mean scores showed SWI students had significantly higher posttest scores than controls (\( p = 0.005 \)).

SWI students were significantly more engaged than controls on both CCTST pre- and posttests. When both semester totals were combined, more SWI than control students took CCTST pre- and posttests, but far more control students than SWI students (9 vs. 1) were disengaged using the engagement measure of the CCTST (14). Eight of the nine disengaged control students were in the fall group, as was the only disengaged SWI student. When engagement results were compared between combined SWI and control groups with \( \chi^2 \) and binomial tests, using control results as expected values, \( p \) approached 0 in both. When analyzed for persistence as above, spring SWI and control labs were not significantly different. When fall and spring groups were combined and compared for persistence as above, SWI and control groups were not significantly different (data not shown).

**DISCUSSION**

Assessing the effects of any teaching method on actual student learning is difficult, and results may vary widely between student groups even when the method is repeated identically, due to varying individual student achievement, motivation, and self-efficacy beliefs, as well as group interactions (31, 32). Many teaching-efficacy studies are also qualitative and are best combined in a holistic approach with quantitative assessment (18). While many teaching-efficacy assessments rely on student self-reporting (12), self-reported surveys often correlate relatively poorly with actual student learning (22, 24, 31, 32, 34). Using a well-validated and quantitative test like the CCTST (17) may help overcome challenges inherent in pedagogic research by yielding more valid and meaningful data than comparing SAT or ACT scores, high school or undergraduate grade-point averages, or self-reported Likert surveys.

This project underscores some common difficulties seen when assessing efficacy of teaching methods. One of the most vexing is the extreme performance variability between sections, which we observed here. Spring SWI and control groups scored slightly higher on CCTST pretests than their fall counterparts, but both underachieved in lecture. Compared with the fall sections, the spring SWI group lecture grade mean (X) declined 0.81 and the spring control X decreased 0.19.

Despite the variability seen, our results clearly support the hypothesis that students engaged in independent research will perform better. When SWI and control students from both terms were combined, SWI students had higher mean lecture grades than control students, approaching significance (\( p = 0.083 \)), and had significantly higher overall grades and higher proportions of students exceeding or equaling their lecture section mean percentage. Combined SWI groups also had significantly fewer lecture grades \( \leq \) D+ than controls, and both fall and spring SWI groups had
significantly higher CCTST posttest total mean scores than controls. Lastly, SWI students had higher CCTST participation and engagement than control students. Unlike other reports (11, 15), we did not observe increased persistence of SWI students, but this may result from limited data, since we tested just two SWI sections over two semesters.

When proportions of spring SWI students who had a final lecture percentage equal to or greater than their lecture final course mean were compared with spring controls, p neared but did not reach significance (p = 0.063 to p = 0.11). We attribute this to a reduced sample number for the spring SWI group caused by the three spring SWI students who did not take lecture, which in turn lowered the power of the analysis to yield non-significant p values. When spring and fall SWI groups were combined and compared with combined controls, significantly more SWI students scored at or above their lecture final average percentage than did controls (p = 0.019, binomial test). Since this combined analysis included supermajorities of SWI and control students from both semesters (~68% of all SWI students and > 74% of all controls), we believe this is a more accurate measure of the SWI’s effect on student lecture grades than those from a single term.

Similarly, when fall and spring lecture mean grades were combined and SWI was compared with controls by the binomial test, results approached, but did not reach, significance (p = 0.083). The reduced power of this analysis from the three SWI students who did not take lecture appears to be the reason for this as well.

The posttest critical thinking results for both fall and spring SWI students suggest using the SWI may help improve student critical thinking. Both SWI cohorts saw posttest critical thinking total average scores increase to 73.6, whereas control students’ posttest total average scores were identical to (fall) or lower than (spring) pretest scores. Since the percentage of students taking CCTST pre- and posttests exceeded 60% of all groups except fall 2014 controls, which had identical pre- and posttest total mean scores, we believe the CCTST total mean scores accurately reflect students’ relative critical thinking skills when they took the tests.

Our results of increased CCTST total mean scores for SWI groups and decreases in one of two control groups are strikingly similar to a previous study, in which CCTST indices improved during Introductory Biology classes taught by community-based inquiry, while the same indices in contemporaneous control classes declined (25). Improvements in critical thinking scores on CCTSTs are quite rare (17), but the discovery-based, unpredictable nature of SWI labs might engage student attention and curiosity more than standard labs, which may have contributed to this.

This study helps validate the premise that the Small World Initiative for Antibiotic Discovery is a viable, and even superior, substitute for current undergraduate labs in nonmajors biology courses in addition to its previously documented positive effects on science students (15, 29). Expanding the SWI into more sections at FAU and elsewhere will allow collection of more data and increase confidence in the relationship between participating in research and increased academic performance, as well as provide a clearer understanding of the effects on persistence. With more participants, the data can also be aggregated by gender, college preparation level, minority status, and first-generation status. If the study can be repeated on a larger scale, converting undergraduate science labs into citizen-science research projects could rapidly improve all student success and align teaching labs more closely with current National Academies of Sciences and American Association for the Advancement of Science recommendations (1). It could also yield valuable biomedical discoveries from areas that funding entities deem too risky or too expensive to support. Most importantly, it may lead to a much greater appreciation and knowledge of science by US society as a whole by providing research experiences to increasing numbers of students who will enter the workforce as young professionals.

ACKNOWLEDGMENTS

We gratefully acknowledge Dr. Rod Murphey, Biological Sciences Chair, and the FAU Foundation, for sending JPC to Yale for SWI training and for buying the CCTSTs. We also thank the Helmsley Foundation and the Howard Hughes Medical Foundation for their support to expand and disseminate this project. We thank Drs. Tiffany Tsang, Jo Handelsman, Todd Kelson, Karen Pelletreau, Brittany Gasper, Erick Warrick, Christine Pribbenow, Jennifer Frederick, and Mr. Simon Hernandez for their extensive help and encouragement. We also thank all other SWI Partners and Partner Instructors, especially Dr. Kristin Butela, Ms. Claudia Lemper, and Dr. Ana Barral, for their many helpful suggestions. Thanks also to our FAU undergraduate researchers for troubleshooting the labs before we used them and for their help during the year. We also thank Dr. Dale Gawlik of FAU and Dr. Patrick Cooper of Lynn University for help with statistical analysis. The authors declare that there are no conflicts of interest.

REFERENCES


15. Handelsman, J. 2014. American Society for Microbiology, President’s forum address, ASM General Meeting, Boston MA, May.