International Journal for the Scholarship of Teaching and Learning

Volume 8 | Number 2 Article 5

January 2014

Introductory biology course reform: A tale of two courses

Michele I. Shuster PhD
New Mexico State University - Main Campus, mshuster@nmsu.edu

Ralph Preszler
New Mexico State University, Main Campus, rpreszle@ad.nmsu.edu

Introductory biology course reform: A tale of two courses

Abstract

Over the past eight years we have undertaken iterative cycles of course reform in two introductory biology courses: Biology 111 and Biology 211. Our revisions of these formerly "traditional" lecture courses have included in-class case studies with and without peer facilitators and peer-facilitated small-group workshops.

Based on analyses of overall pass rates, as well as pass rates by gender and by underrepresented minority (URM) status, we have found that there are differences in the effectiveness of alternative course models in the two courses. In Biol 111, required peer-facilitated workshops improved overall student performance, especially for URM and female students (Preszler, 2009). Here we report that similar workshops were not as successful in Biol 211, but that in-class case studies facilitated by peer instructors have improved student performance and reduced the performance gap. Clearly, what is the "best practice" for one course is not the best practice for the other.

Keywords

undergraduate, peer instructors, introductory biology, majors, course reform

Cover Page Footnote

We thank all of the Biol 211 BioCats and Biol 211 Graduate Teaching Assistants who made the implementation of the revised course models possible. Tonia Lane, Anja Hansen and Claudia Trueblood helped coordinate the hiring of BioCats over the course revision process. We would also like to acknowledge the Biol 211 instructors who have taught in the various course models throughout this process, many of whom also contributed to the development of some of the case studies. Dr. Amy Marion helped develop the original revision model. Dr. C. Brad Shuster helped prepare the digital figures. The NMSU- HHMI undergraduate science education programs, funded by the Howard Hughes Medical Institute and our College of Arts and Sciences provided financial support.

INTRODUCTION

Large-enrollment introductory biology courses continue to be challenging for both students and instructors. Many of these courses are characterized by low pass rates, and are viewed as "gateway" or "barrier" courses (PCAST, 2012). In addition to low overall student performance, there is a consistent pattern of underrepresented minorities (URMs) having lower pass rates than their non-URM peers (*e.g*. Born, Revelle & Pinto, 2002; Haak, HillRisLamber, Pitre & Freeman, 2011; Rath, Peterfreund, Xenos, Bayliss & Carnal, 2007; Villarejo, Barlow, Kogan, Veazey & Sweeney, 2008). This performance gap contributes to the continued underrepresentation of URMs in STEM fields, such that the population of students earning STEM degrees and STEM professionals does not mirror that of the United States (National Academies of Sciences, 2011; National Science Foundation, 2013; Nelson & Brammer, 2010). A lack of diversity in STEM graduates and STEM professionals is detrimental to creativity and continued leadership in STEM fields (Nelson & Brammer, 2010).

Our objective was to investigate the impact of iterative course-based research to guide curriculum reform. We describe several rounds of course reform (using several revised course models) carried out in an effort to improve student success and reduce the performance gap between URMs and non-URMs. If students are able to succeed in their introductory biology classes on their first attempt, they can progress in their major and reduce their time to graduation. However, it is not enough to focus simply on pass rates. It is important to ensure that students who successfully complete our introductory courses are adequately prepared for their subsequent coursework, and that their experiences in introductory courses do not turn them away from the sciences (*e.g*. Tanner & Allen, 2004).

New Mexico State University is the state's land grant institution, it is classified as a RU/H (Research University: high research activity) by the Carnegie Foundation, and is a Hispanic-serving institution. In the fall of 2012, 47% of all students on the main campus were Hispanic, and 55% of the freshman class was Hispanic (New Mexico State University Factbook, Fall 2012). Entering freshman ACT scores for the period included in our study are very stable, and averaged 20.64. In an ANOVA analysis, there are no differences in entering freshmen ACT scores among the course models that we have investigated (ACT data from Fall 2004 and Fall 2013 New Mexico State University Factbooks). We have two introductory biology courses, each of which serves a variety of majors, as well as students who have not yet declared a major. Historically, these courses have had low pass rates. There was also a large disparity in pass rates between URM and non-URM students. With support from the Howard Hughes Medical Institute's (HHMI) Undergraduate Science Education Program and our College of Arts and Sciences, we have transformed each course to improve overall pass rates, and reduce the gap between URM and non-URM students. In addition to improving grades, we aimed to insure that our reforms improved student learning and student interest in science.

A variety of approaches have been described to address student success in introductory STEM courses. Some approaches rely on addressing the preparation of incoming students, providing a preparatory experiences for students prior to their enrollment in the majors introductory course. Such programs include BIOS Boot Camp, University of Washington Biology Fellows Program, and the University of California Berkeley Biology Scholars Program, among others (Buchwitz *et al*., 2012; Dirks & Cunningham, 2006; Matsui, Liu & Kane, 2003; Wichusen & Wichusen, 2007). The preparatory approach has been shown to improve participating students' performance in subsequent introductory biology courses. However, additional benefits can be gained by supplementing or revising introductory courses themselves.

Other approaches involve providing out-of-class learning and studying opportunities for students in the class. While there are many models for these approaches, they generally rely on peer facilitators and focus on study strategies as well as course material. As examples (and not intended as a comprehensive review), these programs include Supplemental Instruction (*e.g*. Rath *et al*., 2007), Triesman-style workshop groups (*e.g.* Born *et al*. 2002; Fullilove & Triesman, 1990), Peer-Led Team Learning (*e.g*. Gafney and Varma-Nelson, 2008; Hockings, DeNagelis &

Frey, 2008), and other forms of study groups (*e.g*. Otero, Finkelstein, McCray & Pollock, 2006; Stanger-Hall, Lang & Maas, 2010). A common feature of these models is that the out-of-class group work occurs as a supplement to the lecture, increasing the time that students spend working on the class material. These models positively impact various student outcomes including overall pass rate and reducing the performance gap between URMs and non-URMs. However, requiring additional meetings outside of regularly scheduled class time can pose a barrier to students who may have extensive work or family commitments. These models also struggle to reach students who do not recognize the effort required to succeed in university-level science courses until they have fallen behind. Voluntary programs do not ensure that a sufficient proportion of students will experience the associated benefits.

Strategies that reach all enrolled students include models of course reform focused on the class itself, typically directed at increasing the amount of active learning and/or frequency of assessment. Among the successful approaches that we have drawn from are strategies to introduce more active and collaborative learning (*e.g*. Armstrong, Chang & Brickman, 2007; Handelsman *et al*., 2004; Knight & Wood, 2005; Tanner, 2009; Walker, Cotner, Baepler & Decker, 2008), to change the nature and frequency of assessment (*e.g*. Casem, 2006; Freeman *et al*., 2007; Freeman, Haak & Wenderoth, 2011; Williams, Aguilar-Roca, Tsai, Wong, Beaupre & O'Dowd, 2011) and to introduce case studies and other problem-based learning to the class (*e.g*. Allen, Duch and Groh, 1996; Gaffney, Richards, Kutusch, Ding & Beichner, 2008; Herreid, 1994).

Some in-class reforms include the use of undergraduate peer instructors. In these cases, the peer instructors facilitate required course activities that take place within the course structure and are integral members of the instructional team (*e.g*. Preszler, 2009; Smith, Stewart, Shields, Hayes-Klosteridis, Robinson & Yuan, 2005). Our successful course reforms have relied on undergraduate peer instructors facilitating integral course activities.

One of the courses that we have successfully transformed is Biol 111, The Natural History of Life (Preszler, 2009). This course serves a variety of science and science-related majors, as well as many students (approximately 23%) who have not yet declared a major. Historically the lecture course met three times a week for 50-minute lectures. As part of our course revision, a mandatory small-group workshop replaced one of the three weekly lectures. While the workshop materials are developed by the course instructor, the workshops themselves are facilitated by undergraduate peer instructors (known as Biology Learning Catalysts, or BioCats). As described by Preszler (2009), the change in course structure was associated with positive student attitudes, as well as large increases in the proportion of A's and B's earned by students, and substantial decreases in the proportion of students earning F's or withdrawing from the course (W's). Even more importantly, while all students appeared to benefit from the course reform, URMs had significantly greater benefits than non-URMs, based on increases in final course grades in comparison to pre-reform semesters (Preszler, 2009).

The focus of this study is on the process of curriculum reform in our other introductory biology course that serves science majors and students with an academic or professional need for biology, Biol 211, Cellular and Organismal Biology. Students in this course are generally first and second year students, representing primarily (but not exclusively) pre-nursing, biology, biochemistry and agriculture majors. This course also had a traditionally very low pass rate (56.5% and 63.8% in two sections prior to any of the course revisions described here) and a performance gap between URMs and non-URMs.

Here we describe several rounds of course reform (revised course models) carried out in an effort to improve student success and reduce the performance gap between URMs and non-URMs in Biol 211. As described below, these course models have included the use of in-class case studies, peer-facilitated and integrated workshops (as described in Preszler, 2009), and peer-facilitated in-class case studies combined with a peer-facilitated Help Desk. After several semesters of implementation of each course model, we evaluated and made changes to the model in order to improve outcomes. Interestingly, the version of the course that worked well in Biol 111 did not achieve comparable results in Biol 211, reinforcing the importance of empirical evaluation, even when implementing “best practices” in a course.

METHODS

Cellular and Organismal Biology (Biol 211), is the second of two introductory courses for biology majors at our institution, but the only introductory biology course taken by biochemistry and pre-nursing students. The lecture course is a 3-credit course, and is separate from the 1-credit Biol 211L laboratory course. Concurrent registration in the lecture and lab is not required, although the vast majority of students enrolled in the lecture do concurrently enroll in the lab. Total enrollments range between approximately 225 and 310 students per semester, and either one or two sections may be offered in each semester. Thus, section sizes range from approximately 125 to 310 students. Course topics include the scientific method, atoms, bonds and molecules, cell structure, enzyme activity, cellular respiration and photosynthesis, molecular genetics (DNA replication, transcription and translation) and some physiology. The course is taught by different instructors, who have the flexibility to spend different amounts of time on individual topics and to adjust the grading scheme for their sections. However, all instructors followed the general course models as described below during this extended course reform process, and one instructor taught 13 of the 25 sections included in this analysis (baseline through three distinct course models).

Control (Baseline)

The control, or baseline, condition was in place from Fall 2003 through Fall 2005 (we are only considering academic year semesters and are excluding summer sessions). During this time, 9 sections were offered by 6 instructors. The 3-credit lecture course met for three 50-minute lectures per week. These were largely traditional lectures, with some activities such as think-pair-share or small group discussion. Beginning in the Fall 2005 semester, clickers were introduced into some sections, adding an element of active learning that was intended to

https://doi.org/10.20429/ijsotl.2014.080205

engage all students, through a small percent of the final grade being earned by scored clicker responses (see Preszler, Dawe, Shuster & Shuster, 2007).

Lecture Cases (LC)

In Spring 2006, in-class case studies were introduced into the lecture. Eight class meetings (one meeting approximately every two weeks) were devoted to working through a case study. In order to ensure that students were accountable for the case studies, case study work product (*e.g*. an in-class assignment or worksheet) accounted for between 30 and 35% of students' grades in this model, and exams included specific questions related to the case studies. The case studies were intended to reinforce the lecture content as well allow students to apply lecture content to novel, interesting and relevant scenarios. Some of the case studies were adapted from published case studies at the National Center for Case Study Teaching in Science (http://sciencecases.lib.buffalo.edu/cs/), and some were written by the instructor. Students were required to complete some form of preparation for the case studies. Typically this was a reading (from the textbook or a website) accompanied by reading questions to be completed before the in-class case study. Some of the questions were more specific to the case, and were essential to work through the case (*e.g*. cancer statistics or nutrition information from specific foods).

The instructor acted as a facilitator during each case study, ensuring that students kept on task and on time. A graduate student teaching assistant also helped facilitate the case study sessions, by circulating through the lecture room with the instructor and helping student groups that had questions. Both the instructor and teaching assistant were careful not to provide direct answers during these sessions, but did provide scaffolds to help students break their questions down into more manageable (and answerable) questions.

In addition to the case studies, clickers continued to be a component of the LC model and grading scheme, accounting for 15% of the final grade.

Workshops (WRK)

We decided to capitalize on the success of the case study approach by having students work in small-group, peer-facilitated workshops (see Preszler, 2009 for a complete description of the our workshop and peer instructor model). In this model, students met in the large lecture for two 50-minute meetings per week. Instead of a third lecture meeting, each student registered for and attended a mandatory 65-minute workshop. Each workshop enrolled up to 24 students and was facilitated by a BioCat (undergraduate peer instructor), who also attended every lecture in the course. The workshop activities were very similar to the in-class case studies in design and intent. However, the workshop format and length allowed for more student interactions. Each student group had a large whiteboard, allowing them to present their work to other groups. The instructors prepared all the workshop materials, and spent approximately one hour per week training the BioCats on the up-coming workshop. The BioCats suggested modifications during each training session and provided feedback on the previous week's workshop. The BioCats also graded the students' workshop assignments.

The workshops contributed to between approximately 20% and 30% of the course grade, depending on the semester. The workshop model also included interactive lectures with clickers and a variety of forms of in-class student talk (Tanner, 2009). The workshop model was implemented for four semesters.

Lecture Cases with BioCats (LCBC)

Due to disappointing outcomes with the WRK model (see results below), we decided to revise the course. It was clear that the LC model had been more successful than the WRK model in Biol 211. We thus decided to build on the prior experience, enhancing it with the addition of BioCats.

In the current LCBC model, there are two 75-minute lectures each week. The lectures are interactive, with clicker questions (students are encouraged to discuss the questions with their neighbors), think-pair-shares and student-generated questions. Approximately once every two weeks, one of the lecture sessions is devoted to an in-class case study, facilitated by the Instructor and the BioCats (eight in a single large section, or four BioCats in each of two smaller sections). As in the WRK model, the BioCats attend every lecture, and grade student case study assignments. As in the LC model, students complete a preparatory assignment before each in-class case study. A series of clicker questions based on the prep assignment is used in an effort to ensure student accountability for the prep assignment.

One feature of the LCBC model that extends the impact of BioCats is a Bio Help Desk. Each BioCat schedules three hours of Bio Help desk each week, resulting in approximately 24 hours of Biol 211 Help Desk each week. BioCats at the Help Desk are available to help students with questions about the course material. The BioCats help the students by breaking their questions down into smaller steps, asking students to draw a process on the whiteboard available at the Help Desk, and/or asking the students to explain their answers. In a recent (typical) semester, 21% of course students signed in at Help Desk at least once, and 6% of students signed in more than two times during the semester.

Table 1 summarizes each of the course models described here. We are reporting on 16 academic year semesters from Fall 2003 to Spring 2011.

Table 1: Summary of the Different Course Models

Course Model	Semesters/Instructors*	Description
Traditional (CNTRL) (1,180 students)	Fall 2003-Fall 2005 (5 semesters/9 sections, 6 instructors; A (2 sec), B, C (3 sec), D, E & F)	Interactive lectures with clickers
Lecture Cases (LC) (432 students)	Spring and Fall 2006 (2 semesters/3 sections, 2 instructors; C (2 sec.) & E)	Interactive lecture with clickers; In-class case studies (~8 per semester) facilitated by single graduate teaching assistant and Instructor

Workshops (WRK) (905 students)	Spring 2007- Fall 2008 (4 semesters/4 sections; 2 instructors; C (3 sec) & E)	Two interactive 50 min lectures with clickers, one 65-minute workshop-facilitated by BioCats
Lecture Cases with BioCats (LCBC) (1,444 students)	Spring 2009- Spring 2011 (5 semesters/9 sections; 3 instructors; C (5 sec), G (3 sec) & H)	Two 75-minute lectures per week; interactive plus full-period in-class studies every 2 weeks, led by Instructor & facilitated by BioCats. BioCats facilitate Help Desk.

*Individual instructors are designated with a letter (A-H), and listed in each model. For instructors that taught more than one section in each model, the number of sections is indicated.

Assessment Overview

In order to determine whether the course models were meeting our goals of increasing student success, we examined overall course grades in each course model across all course instructors. We additionally examined grades based on gender and ethnicity, to see if any course model was differentially impacting specific groups of students in the course. We recognize that tracking course grades can be confounded by grading schemes, particularly the proportion of points associated with exams. As discussed below, all three revised models added points associated with the main intervention (relative to control/no intervention). Among all the revised models, the percent of points associated with each feature (e.g. in-class case studies and workshops) ranged from a low of 20.5% (one WRK semester) to a high of 35% (the first LC semester), typically hovering around 30%.

We used specific questions from student evaluations of the course in order to determine student opinions of each course model. As many factors influence student opinions and responses on student evaluations, we are only reporting the student evaluations for one instructor ("C") who taught a large number of the sections in each of the revised models (two of the LC sections, three of the WRK sections, and five of the LCBC sections).

We also monitored student performance on multiple-choice exam questions on a traditionally challenging topic (cellular respiration), to see if overall improvements in student grades were paralleled by improvements in performance on a specific course topic. Again, to reduce sources of variability in this more fine-scale analysis, we are only reporting exam performance from the same single instructor.

Course Grades

Relationships between course grades and course model, gender and ethnicity were evaluated using two-way and three-way contingency table analyses. In all cases, if the probability associated with the Pearson χ^2 was <0.01, we concluded that the variable(s) in question had a significant impact on student grades.

We used two-way contingency tables to look at the impact of course model on course grades- specifically, whether the distribution of grades differed in the different course models. We also used two-way contingency tables to determine whether grade distributions differed between females and males, and whether grade distributions differed between URMs and non-URMs. URMs are students who self-identified as being African-American, Latino or Native American, and non-URMs are students who have self-identified as Asian American or Caucasian. Students who chose not to identify a race or ethnicity, or who selected "other" during the institutional application process were not included in the ethnicity analysis.

The three-way contingency table analyses were used to determine the impact of the different course models on the relationship between gender and grades (did females and males respond similarly to the different course models?) and ethnicity and grades (did URM and non-URM students respond similarly to the different course models?).

In addition to evaluating grade distributions, we also examined percent changes in grades, relative to the control. In these cases, we first calculated the percent of students earning each letter grade in each case (*e.g.* the percent A's, B's, C's, D's, F's and W's earned by URMs in each course model). We then subtracted the % of each grade in the control from the % of each grade in a given model (*e.g*. the % of A's earned by URMs in the control was subtracted from the % of A's earned by URMs in the LC model). This difference was then expressed as a % of the value in the control. As a hypothetical example, if the % of A's earned in the control was 14%, and the % of A's earned in a particular model was 20%, the difference is 6%, which is a 42.9% improvement in A's relative to the 14% in the control.

Scores on cellular respiration exam questions

In order to determine if observed improvement in course grades was accompanied by an improvement in understanding of a specific topic, rather than an artifact of changing course grading schemes, midterm and final exam questions pertaining to this topic were analyzed from certain sections taught by a single instructor ("C"). Each question was evaluated for the percentage of students that answered it correctly within a class section. For some of the older semesters, either no data was available, or only partial data was available (*e.g*. questions from only one version of the midterm, representing only a subset of students). Table 2 shows the data available for this analysis. In addition to plotting the averages for each course model, we used an ANOVA to investigate whether there were significant differences (p value < 0.05) in cellular respiration exam question scores in the different course models.

Table 2: Cellular Respiration Exam Questions

Semester	Course Model	# CR Exam Questions	# Students
Sp03	Control	16	33
Sp04	Control	10	54

Fa04	Control	10	27
Fa05	Control	13	29
Fa06	LC	14	105
Sp07	WRK	16	188
Sp09	LCBC	37	152
Fa09	LCBC	30	127
Sp10	LCBC	32	273

Student Evaluations

As one instructor taught a substantial number of the sections in each revised format, we examined course evaluation data for that instructor. Focusing on a single instructor who taught in all four versions of the course allowed us to compare student evaluations of the four course models without confounding the comparisons with instructor effects. Specifically, we focused on how students responded to three questions about the course format: whether the specific format (LC, WRK, LCBC) made them more interested in the course content, whether the specific format (LC, WRK, LCBC) helped them understand the content, and whether the specific format (LC, WRK) was a positive addition to the course. Students were also asked about their current interest in biology. These questions were embedded on the anonymous end-of-semester student evaluations of the course. Students responded on a 5-point Likert scale (strongly agree, agree, neutral/no opinion, disagree and strongly disagree). The percent of students selecting each response was calculated for each section/semester. These percentages were averaged for each course model, to obtain an overall student evaluation of each course model.

This research was reviewed and approved by the institutional IRB (protocol # 354).

RESULTS

Grades and Course Model

The distributions of grades differed significantly between the four different course models (Pearson chi-squared $p<0.001$). The distributions of the percentages of each letter grade are shown in Table 3. The percent change of each letter grade relative to the control is shown in Figure 1. The significance of the differences appear to be driven by large increases in B's relative to the control, and a decrease in F's and W's relative to the control (Figure 1). These trends are generally consistent for each of the three revised course models.

Table 3: Distribution of the Percentages of Letter Grades in Each Course Model

	"A"	"B"	"C"	"D"	"F"	"W"
Control	13.31	19.15	24.83	12.37	18.39	11.95
LC	15.28	29.63	22.45	10.42	15.28	6.94
WK	15.91	24.53	22.32	13.70	15.25	8.29
LCBC	16.21	26.94	23.48	14.20	11.08	8.10

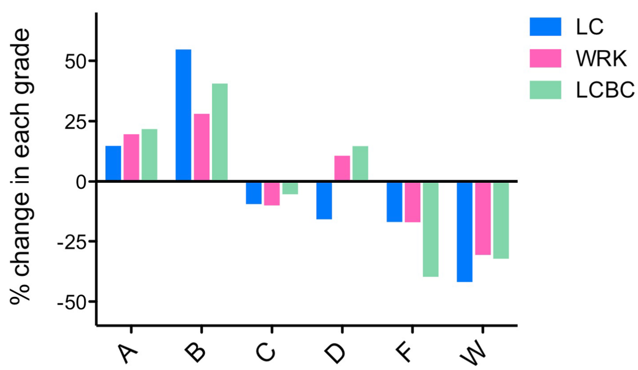

Figure 1: Changes in Grades with Each Course Model

Grades and Gender

There were no significant differences in the grade distributions of males and females, when pooled from Fall 2003 through Spring 2011 (*i.e*. across all the course models) (Pearson Chi-squared p=0.75; Table 4).

Table 4: Overall Course Grades (Fall 2003- Spring 2011) for Males and Females

	"A"	"B"	"C"	"D"	"F"	"W"
Males	15.91	24.35	24.03	12.34	13.88	9.50
Females	14.84	24.37	23.27	13.49	15.02	9.01

While males and females did not have different distributions of grades when averaged across all four course models, males and females did respond differently to changes in course models (3 way contingency table, Pearson chi-squared $p<0.001$). The percent changes (relative to control) are shown in Figure 2. In comparison to the control semester, female's percent increase in A's and B's was highest in the LC model, and the LC model resulted in the largest reduction of W's (course withdrawals) for females. In contrast, males showed no increase in A's with the LC model, large increases in A's and B's with the LCBC model and concurrent reductions in D, F, and W's with the LCBC model. In general, females performed best with the LC model, while males' performance was highest with the LCBC model.

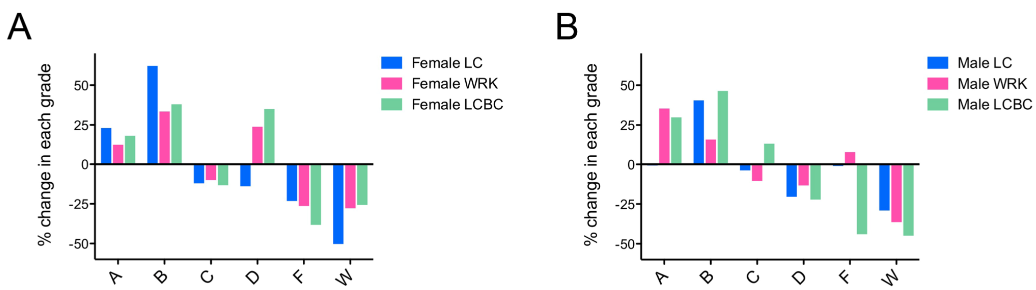

Figure 2: Changes in Grades for Females and Males in Different Course Models a. Female grades b. Male grades

Grades and URM Status

When looking at the overall distribution of grades pooled from Fall 2003 through Spring 2011 (*i.e*. across all the course models), URMs and non-URMS performed significantly differently from one another (Pearson Chi-squared p<0.001) (Table 5). In this case, non-URMs are doing significantly better than their URM peers. This is particularly evident in the percent of students earning A's and the percent of students earning D's and F's.

Table 5: Overall Course Grades (Fall 2003- Spring 2011) for URMs and Non-URMs

	"A"	"B"	"C"	"D"	"F"	"W"
URMs	9.63	21.94	24.42	16.30	18.05	9.68
Non-URMs	21.67	27.87	22.27	9.33	10.73	8.13

URM and non-URM students responded differently to the sequence of course models (3-way contingency table, Pearson chi-squared p<0.001). A striking finding is that URM students in the LCBC model showed the most substantial percent reduction in F's and W's relative to the control (Figure 3). In terms of trends of overall "best grades", URM students seemed to do best with the LCBC model, followed by LC, which was better than WRK, which in turn was better than the control. In contrast, non-URM students seemed to do best with LC, followed by LCBC and WRK (essentially the same), and did the worst in the control semesters.

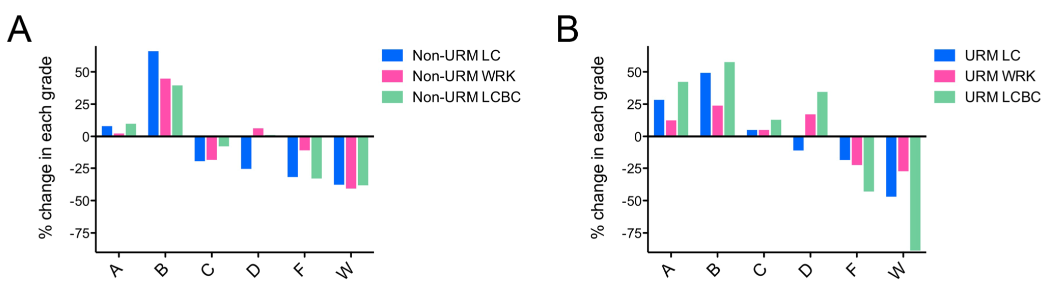

Figure 3: Changes in Grades for Non-URM and URM Students in Different Course Models. a. Non-URM Students b. URM Students

Cellular Respiration Exam Performance

As cellular respiration is a challenging topic for students in this course, we have monitored aggregate exam scores for questions on the topic of cellular respiration, all from sections of a single instructor. In semesters that incorporated case studies, some of the exam questions were directly related to the case studies, and some were extensions of the case studies (generally transfer to a novel scenario).

The overall percent correct on the cellular respiration exam questions in the semesters for which we have data are shown in Figure 4. As can be seen, cellular respiration scores remained relatively stable between the control, LC and WRK models, then appear to increase with the LCBC models. A one-way ANOVA showed that the scores differed significantly across these models ($p=0.03$), presumably due to the increase in scores with the LCBC model.

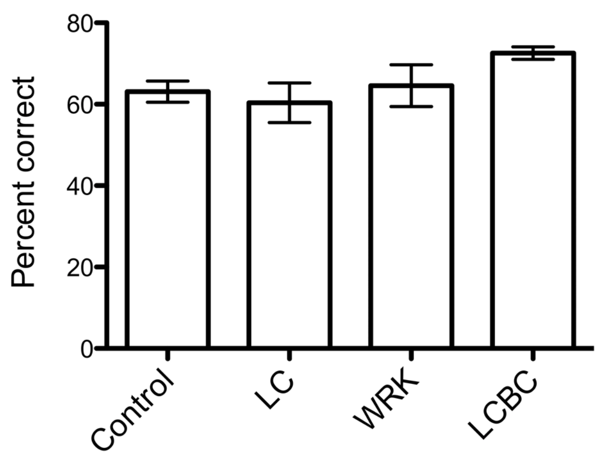

Figure 4: Cellular Respiration Exam Scores (Error bars represent standard error)

Student Evaluations

As one instructor taught a substantial number of the sections in each format, we examined course evaluation data for that instructor for common items asked in each section and semester. These data represent averages for one LC semester

(Fa06), 5 sections in 3 semesters of WRK, and 5 sections in 4 semesters of LCBC.

Students responded to questions about whether the specific format (LC, WRK, LCBC) made them more interested in the course content, whether the specific format (LC, WRK, LCBC) helped them learn the course content, and whether they thought that the specific format (LC, WRK) was a valuable addition to the course (for LC and WRK models only). The positive responses (strongly agree and agree) have been combined, as have the negative responses (disagree and strongly disagree).

Students in the LCBC semesters were more positive than students in other models about the impact of the in-class activities on their interest in the course material. Students in the WRK semesters had the lowest opinions about the ability of the workshops to enhance their interest in the course material (Table 6).

Table 6. Student Perceptions of the Impact of Each Component on their Interest*.

Response	SA/A	N	D/SD
LC (Avg. % of responses)	59	24	17
WRK (Avg. % of responses ± SE)	54.4 ± 3.5	24.8 ± 2.7	22.6 ± 3.3
LCBC (Avg. % of responses ± SE)	67.4 ± 4.7	24.9 ± 3.1	7.6 ± 1.6

SA/A: Strongly agree/agree; N: Neutral/no opinion; D/SD: Disagree/strongly disagree

*Students responded to the item "____________ *made me more interested",* where the blank was in-class activities in LC and LCBC semesters and workshops in WK semesters.

Students' perception of the impact of course activities on their understanding of course material was substantially higher in the LCBC semesters than in the other models. Students in LC and WK semesters had similar opinions about how in-class activities and workshops contributed to their understanding of the course material (Table 7).

Table 7. Student Perceptions of the Contribution of Each Component to their Understanding*.

Response	SA/A	N	D/SD
LC (Avg. % of responses)	66	16	18
WRK (Avg. % of responses ± SE)	62.3 ± 2.6	15.7 ± 1.3	21.3 ± 2.4
LCBC (Avg. % of responses ± SE)	82.2 ± 3.0	11.8 ± 1.7	6.0 ± 1.7

SA/A: Strongly agree/agree; N: Neutral/no opinion; D/SD: Disagree/strongly disagree
*Students responded to the item "___________ *helped me understand",* where the blank was in-class activities in LC and LCBC semesters and workshops in WK semesters

Students in LC and WRK semesters were asked their opinions about whether in-class activities and workshops (respectively) were positive additions to the course. In both cases, the majority of students agreed that these components were positive additions. 65% and 64.8% ± 3.8 of LC and WRK students respectively strongly agreed or agreed that the in-class case studies and workshops were a positive addition to the course.

Students were also asked to rate their interest in biology at the end of the course, relative to the start of the course. Students responded on a 5-point scale (much higher, somewhat higher, about the same, somewhat lower, much lower). Responses have been collapsed into three categories to capture those who were more interested, had same the interest, and were less interested at the end of the course (Table 8).

While the majority of students in all sections indicated that they had more interest in biology at the end of the course relative to the start of the course, students in the workshop sections expressed the lowest amount of enhanced interest (only 56% on average were more interested in biology at the end of the course), and also the highest loss of interest (nearly 12% had less interest in biology at the end of the course).

Table 8: Student responses to the item "Compared to when I started this course, my interest in biology now is _________."

Response	Much/Somewhat higher	About the same	Much/somewhat lower
LC (Avg. % of responses)	68	24	5
WRK (Avg. % of responses ± SE)	56 ± 2.5	32.4 ± 2.6	11.7 ± 1.5
LCBC (Avg. % of responses ± SE)	64.8 ± 3.7	28 ± 3.6	6.9 ± 1.0

DISCUSSION

We have monitored overall course grades across four course models, as well as how grades of males and females and URMs and non-URMs respond to the different course models. We have also monitored exam performance on a discrete and challenging topic, and student opinions of the value of different course models at promoting interest and understanding of the course material. By triangulating these different data sources, we have shown that the models were not equally effective, and that groups of students responded differently to course models. However, based on the available data, the LCBC appears to be the best model for Biol 211 at our institution, while the WRK model was the least successful of the revised models. This is surprising, given the success of the WRK model in Biol 111 at our institution (Preszler, 2009). We can only speculate on the reasons why groups of students in Biol 211 responded differently to course models and why the evidence demonstrated that one approach (WRK) was not equally successful in two introductory biology courses at the same institution.

The LC and LCBC models involve instructor-facilitated in-class case studies. In the LC model, one instructor and one graduate teaching assistant facilitated in-class activities for between approximately 125 and 250 students in a section- a facilitator to student ratio ranging from 1/60 to 1/125. During the in-class case study sessions in the LC model, the instructor and graduate TA were kept continuously busy, and were not able to get to every group that had a question at any given moment. It is possible that some students did not get their questions addressed, particularly shy students. They could have thus left the class meeting with an incomplete understanding of the case study and how it related to course content. This idea is supported by our direct measures of learning of cellular respiration (lower in LC semesters than LCBC semesters, Figure 4), and student opinions of how helpful the case study activities were in helping them understand the course material (higher for LCBC than LC) (Table 7).

In the LCBC model, several BioCats (approximately one BioCat per 40 students) were added to the facilitation team. This increased the facilitator to student ratio to approximately 1/30. In addition to increasing the number of facilitators available to help students during the in-class case study activities, the BioCats may provide a more approachable source of help. Additionally, the collaboration between the Instructor, Graduate Teaching Assistant, and the BioCats is mutually beneficial. It provides the BioCats with "backup" for complicated questions regarding content, and provides the Instructor with a better sense of what students may be struggling with due to immediate feedback from BioCats, who collectively interact with far more students than the Instructor. The benefits that emerge from this expanded instructional team may contribute to the enhanced success of the LCBC model over the LC model.

While the WRK model shares BioCats with the LCBC model, the WRK model was not as successful in Biol 211. There are many possible reasons for this, one is that the synergy between the BioCats and the Instructor is lost when the BioCats are the sole facilitators of case-study activities in the workshop setting. In workshops, the BioCats do not have the content back-up provided by the Instructor, and the instructor does not have the immediate feedback from BioCats during the associated lecture.

We were surprised that males and URMs experienced the greatest benefits with the same model (LCBC), and that females and non-URMs experienced the greatest benefits with a different model (LC). While we have not tested a mechanism underlying this result, we have generated an untested mechanistic hypothesis based on our observations. Students who feel marginalized or lack confidence due to their membership in a group that is under-represented in the classroom (males) or a group that is under-performing (URM) may be less likely to put up their hands or seek assistance in the LC model, but more likely to seek assistance from a BioCat in the LCBC model. At ~ 30% of the students, males are numerically underrepresented in this course. While URM students are not numerically underrepresented in this study, our data suggest that URM students are less prepared (based on the performance gap in control semesters), and may have limited self-confidence in asking questions directly to an Instructor, relative to a BioCat.

A second possible explanation of the pattern of males and URM students performing best in the LCBC model, while females and non-URM students performed best in the LC model, is associated with differences in the diversity of LC and LCBC instructional teams. In the LC model both the instructor and graduate TA were female. The BioCats bring some males to the instructional team in the LCBC model. While we have not made systematic observations, male students may be more likely to seek help from a male BioCat than a female instructor. In addition to the gender diversity introduced with BioCats, the BioCats bring ethnic and racial diversity to the instructional team. Overall, the LCBC instructional team has had a greater amount of gender, ethnic and racial diversity relative to the LC model. While these potential explanations have not yet been tested, this unexpected contrast between gender- and ethnicity-based responses to changes in course models highlights the complexity of the dynamic between students and instructional teams. It also highlights the need to rely on evidence rather than preconceptions when considering the effectiveness of alternative instructional models.

The LCBC model also includes a BioCat-facilitated Help Desk. We do not have sufficiently detailed records of Help Desk visits to know if specific groups of students are taking advantage of this resource more than others. In general, the Help Desk is woefully underutilized, except for the few days before each exam. The informal records that we keep suggest that fewer than 10% of the students in the class visit the Help Desk on a regular basis (more than two times during the semester). In a recent semester, only 6% of the students signed in at the Help Desk more than two times during the semester. It thus seems unlikely that Help Desk itself can explain the differential benefits of the LCBC model over the LC model.

We are still left with the surprising finding that the WRK model was far more effective in Biol 111 than in Biol 211 based on the magnitude of changes in grades from the control (Figure 3 and Preszler, 2009). There is an abundance of literature showing that small, peer facilitated groups enhance student performance, particularly for URM students (*e.g*. Born *et al*., 2002; Fullilove & Triesman, 1990; Otero *et al*., 2006; Rath *et al*., 2007). We speculate that the different student populations in these two courses may be important in the differential success of the WRK model. In Biol 111, there is a higher proportion of first year students, who are new to the University and less academically mature. Such students may benefit from the small workshop environment to discuss with peers and BioCats not only the course content, but also challenges they are facing as new University students. The BioCat workshop facilitators may serve a variety of roles in Biol 111, including assisting with course content, but also modeling the traits of a successful student. This latter role model may be very important for the beginning students in Biol 111. The students in Biol 211 are not necessarily new to the University, and as the content becomes more sophisticated in Biol 211, it may be that the BioCats are most effective as members of an instructional team that works best in a large classroom setting, where there are many resources to help overcome difficulties with the content.

Overall, the LCBC model seems to promote student learning and engender positive student attitudes over the other revised models in Biol 211. Superficially, the LCBC model may appear to be more resource-intensive, but the BioCat resources are far less expensive than graduate assistants. As noted by Otero *et al*. (2006), course reform in the absence of undergraduate learning assistants was not as effective as course reform taking advantage of undergraduate learning assistants. This is similar to our findings that LCBC was generally superior to LC. And similar to Otero *et al*. (2006), at $1500 (each) per semester, even 8 BioCats are far less expensive than 2 graduate assistants. While the LC model relied on a single graduate assistant, this would not have been sustainable, from the perspective of the single graduate assistant and their higher workload relative to other teaching assistants in the department. When considering workload, facilitator to student ratios, and effectiveness, several BioCats are less expensive and equally or more effective than relying on one or two graduate teaching assistants to take on the same roles.

As an Hispanic-serving Institution in a state with a majority minority population (U.S. Census Bureau, 2013), and with the calls to increase the STEM graduation rates of URMs at a national level (e.g. PCAST, 2012), the data support the LCBC model to meet national goals of diversifying the pool of STEM graduates (National Academies of Sciences, 2011; National Science Foundation, 2011; Nelson & Brammer, 2010). Furthermore, this model does not disadvantage any group of students relative to the control. Finally, when considering student opinions as part of our decision-making process, the LCBC model appears to have the most positive impact on student attitudes.

Limitations

We recognize that tracking course grades can be confounded by grading schemes, particularly the proportion of points associated with exams. While overall improvements in grades relative to the control may be associated with changes in the grading schemes relative to the control (*e.g*. the introduction of points associated with in-class case studies and workshops and a corresponding reduction in points associated with exams), all three revised models had points associated with the main intervention. The control sections had the highest percentage of points from exams (77% and 81% in two representative semesters). Of the revised course models, the LC model had the lowest percentage of points from exams (50% and 55% in two LC semesters), followed by WRK (55%-62.5%) and then closely by LCBC (59.7%-65.5%). Among all the revised models, the percent of points associated with each feature (e.g. in-class case studies and workshops) ranged from a low of 20.5% (one WRK semester) to a high of 35% (the first LC semester), typically hovering around 30%. Thus, while the grading model for each revised course differed from the control, there was less variability between the individual course models.

While we don't have a direct measure of possible changes in student preparedness over time during our study, an analysis of ACT scores of entering freshmen suggests that there is no change in preparedness over the course models ($p=0.86$).

When evaluating changes in grade distributions in males and females, or in URMs and non-URMs among course models, we are comparing grades generated by the same grading schemes, factoring out bias associated with specific grading schemes. Additionally by tracking performance on a specific topic, we have shown that performance on that topic has generally followed the trend in course grades, indicating that improved learning is associated with the improved grades.

We acknowledge that one instructor ("C") taught a large number of the sections in this study, and the results may be influenced by the style of teaching that works best for this instructor. However, despite these limitations, the grade distributions show robust and consistent trends across several semesters and faculty members.

Conclusion

While workshops have proven to be successful in one introductory course (Preszler, 2009), they were not as successful in another introductory course at the same institution. One of the lessons that we have learned from this experience is that it is critical to track course-specific outcomes in order to determine "best practice" in the context of a particular course at a particular institution. Instructors should be prepared to respond to assessment data in order to continuously adapt approaches to enhance student success.

https://doi.org/10.20429/ijsotl.2014.080205

REFERENCES

Allen, D.E., Duch, B.J., & Groh, S.E. (1996). The power of problem-based learning in teaching introductory science course. New Directions for Teaching and Learning 68: 43-52.

Armstrong, N., Chang, S-M., & Brickman, M. (2007). Cooperative learning in industrial-sized biology classes. CBE Life Sci Educ 6: 163-171.

Born, W.K., Revelle, W., & Pinto, L.H. (2002). Improving biology performance with workshop groups. J Sci Educ Tech 11: 347-365.

Buchwitz, B.J., Beyer, C.H., Peterson, J.E., Pitre, E., Lalic, N., Sampson, P.D., & Wakimoto, B.T. (2012). Facilitating long-term changes in student approaches to learning science. CBE Life Sci Educ 11: 273-282.

Casem, M.L. (2006). Active learning is not enough. J Coll Sci Teach 35: 52-57.

Dirks, C., & Cunningham, M. (2006). Enhancing diversity in science: Is teaching science process skills the answer? CBE Life Sci Educ 5: 218-226.

Freeman, S., O'Connor, E., Parks, J.W., Cunningham, M., Hurley, D., Haak, D., Dirks, C., & Wenderoth, M.P. (2007). Prescribed active learning increases performance in introductory biology. CBE Life Sci Educ 6: 132-139.

Freeman, S., Haak, D., & Wenderoth, M.P. (2011). Increased course structure improves performance in introductory biology. CBE Life Sci Educ 10: 175-186.

Fullilove, R.E., & Treisman, P.U. (1990). Mathematics achievement among African American undergraduates at the University of California, Berkekey: An evaluation of the mathematics workshop program. The Journal of Negro Education 59: 463-478.

Gaffney, J.D.H., Richards, E., Kutusch, M.B., Ding, L., & Beichner, R.J. (2008). Scaling up education reform. J Coll Sci Teach 37(5): 48-53.

Gafney, L., & Varma-Nelson, P. (2008). Peer-led team learning. Evaluation, dissemination, and institutionalization of a college level initiative. Innovations in Science Education and Technology 16 (Ed. Cohen, K). Springer Science.

Haak, D.C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. Science 332: 1213-1216.

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., Gentile, J., Lauffer, S., Stewart, J., Tilghman, S.M., & Wood, W.B. (2004). Scientific Teaching. Science 304: 521-522.

Herried, C.F. (1994) Case studies in science- a novel method of science education. J Coll Sci Teach 23: 221-229.

Hockings, S.C., DeNagelis, K.J., & Frey, R.F. (2008). Peer-led team learning in general chemistry: Implementation and evaluation. J Chem Educ 85: 990-996.

Knight, J.K., & Wood, W..B (2005). Teaching more by lecturing less. CBE Life Sci Educ 4: 298-310.

Matusi, J., Liu, R., & Kane, C. (2003). Evaluating a science diversity program at UC Berkeley: More questions than answers. CBE Life Sci Educ 2: 117-121.

National Academies of Sciences. 2011. *Expanding Underrepresented Minority Representation: America's*

Science and Technology Talent at the Crossroads. Washington, DC: The National Academies Press.

National Science Foundation, National Center for Science and Engineering Statistics. 2013. *Women, Minorities and Persons with Disabilities in Science and Engineering*, 2013. Special Report NSF 13-304. Arlington, VA. Available at http://www.nsf.gov/statistics/wmpd/

Nelson, D.J. & Brammer, C.N. (2010). A National Analysis of Minorities in Science and Engineering Faculties at Research Universities. 2nd Edition. http://faculty-staff.ou.edu/N/Donna.J.Nelson-1/diversity/Faculty_Tables_FY07/07Report.pdf *(accessed December 3, 2013).*

New Mexico State Universtiy Fall 2012, 2004, 2013 Factbooks. (http://oia.nmsu.edu/data-reports/oiareports/factbooks/) (accessed July 19, 2014).

Otero, V., Finkelstein, N., McCray, R., & Pollock, S. (2006). Who is responsible for preparing science teachers? Science 313: 445-446.

PCAST STEM Undergraduate Education Working Group, Gates SJ, Handelsman J, Lepage GP, Mirkin C. (co-chairs) (2012). Engage to Excel: Producing one million additional college graduates with degrees in science, technology, engineering and mathematics. President's Council of Advisors on Science and Technology. http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_feb.pdf (accessed December 3, 2013)

Preszler, R. (2009). Replacing lecture with peer-led workshops improves student learning. CBE Life Sci Educ. 8:182-192.

Preszler, R., Dawe, A., Shuster, C., & Shuster M. (2007). Clickers: assessing their effects on student attitudes and student learning in biology lectures. CBE Life Sci Educ. 6:29-41.

Rath, K.A., Peterfreund, A.R., Xenos, S.P., Bayliss, F., & Carnal, N. (2007). Supplemental instruction in introductory biology I: Enhancing the performance and retention of underrepresented minority students. CBE Life Sci Educ 6: 203-216.

Smith, A.C., Stewart, R., Shields, P., Hayes-Klosteridis, J., Robinson, P., & Yuan, R. (2005). Introductory biology course: A framework to support active leaning in large enrollment introductory science courses. CBE Life Sci Educ 4: 143-156.

Stanger-Hall, K.F., Lang, S., & Maas, M. (2010). Facilitating learning in large lecture classes: Testing the "Teaching Team" approach to peer learning. CBE Life Sci Educ 9: 498-503.

Tanner, K.D. (2009). Talking to learn: Why biology students should be talking in classrooms and how to make it happen. CBE Life Sci Educ 8: 89-94.

Tanner, K. and Allen, D. (2004). Approaches to biology teaching and learning: Learning styles and the problem of instructional selection- engaging all students in science courses. CBE Life Sci Educ 3: 197-201.

U.S. Census Bureau: State and County Quick Facts. Last revised June 27, 2013. http://quickfacts.census.gov/qfd/states/35000.html (accessed December 3, 2013).

Villarejo, M., Barlow, A.E.L., Kogan, D., Veazey, B.D., & Sweeney, J.K. (2008). Encouraging minority undergraduates to choose science careers: Career paths survey results. CBE Life Sci Educ 7: 394-409.

Walker, J.D., Cotner, S.H., Baepler, P.M., Decker, M.D. (2008). A delicate balance: integrating active learning into a large lecture course. CBE Life Sci Educ 7: 361-367.

Williams, A.E., Aguilar-Roca, N.M., Tsai, M., Wong, M., Beaupre, M.M., & O'Dowd, D.K. (2011). Assessment of learning gains associated with independent exam analysis in introductory biology. CBE Life Sci Educ 11: 346-356.

Wischusen, S.M., Wischusen, E.W. (2007). Biology intensive orientation for students (BIOS): A biology "boot camp". CBE Life Sci Educ 6: 172-178.