Introductory science, math, and engineering courses often have problems related to student engagement, achievement, and course completion. To begin examining these issues in greater depth, this pilot study compared student engagement, achievement, and course completion in a small and large section of an introductory biology class. Results based on t-tests indicate that students in the small section of the course were more engaged and had higher achievement scores than students in the large section. Additionally, students in the small section were more likely to complete the course than students in the large section. Although further research is needed, reduced class size has the potential to be an effective strategy for improving student engagement, achievement, and course completion in an introductory biology course that has too often been a gatekeeper instead of a gateway for students who intend to be science, technology, engineering, and math majors.

Colleges and universities in the United States face the challenge of producing the science, technology, engineering, and math (STEM) professionals who are critical to the world’s future. The National Research Council (2009) reported that the “United States and the world face serious societal challenges in the area of food, environment, energy, and health . . . [and] innovations in biology can lead to sustainable solutions for all of these challenges” (p. 87). At the same time, Hanauer and Bauerle (2012) indicated that “undergraduate science education in the United States is not as effective as it needs to be in translating student interest in science into optimal preparation either to enter the science workforce or to participate as literate citizens in an increasingly global society” (para. 2). Fewer than half of those students who enter college intending to earn an undergraduate degree in a STEM field actually complete the degree within 6 years (Hurtado, 2015). Researchers have blamed large class sizes in introductory classes, lack of engaging pedagogy, and teaching methods that encourage passive learning for students’ decisions to abandon STEM degree programs (Garcia, Gasiewski, & Hurtado, 2011). Despite these known issues, many degree programs and introductory courses continue to rely on large classes and teaching methods that encourage passive learning.

Courses in STEM fields cover a wide array of disciplines. This study focuses on an introductory biology course that is taken as a general education course as well as a prerequisite to professional degrees in a health-related field such as pharmacy, dental hygiene, dentistry, and premed. Students typically take introductory biology courses in their first year of college, which means they are adjusting to a new educational environment as well as to large lecture-style introductory biology courses. Often these courses are fast-paced lectures containing a copious amount of information that students must learn or memorize. For many students, the course experiences can be frustrating and discouraging (Seymour & Hewitt, 1997). Introductory biology courses are intended to be gateways to rewarding, engaging majors and careers. However, for too many students, the courses serve as gatekeepers from their intended field of study instead of gateways to their major or expected career. Although there is literature that highlights best practices for student completion and success in STEM courses (e.g., Hanauer & Bauerle, 2012; Seymour & Hewitt, 1997), change related to STEM course delivery has been elusive. To explore these issues further, we conducted a pilot research study.
study to examine the potential relationship between class size and student engagement, achievement, and completion of an introductory biology course.

**Class size and student achievement**

Although there is a large body of research on correlations between class size and student-related outcomes, results are mixed. There is little commonality of constructs and variables across the literature. Research on the effects of class size on student learning and achievement in K–12 education addresses a smaller range of class sizes than does the research concerning class size in higher education, with large classes in K–12 often well under 40 students (Chingos, 2013; Shin & Chung, 2009). Because of the large range of definitions of class size across the literature, this study relied on class-size research conducted in higher education settings. In higher education, large classes often range from those with more than 70 students to those with hundreds of students, whereas small classes tend to have enrollments in the 20–30 range. Although the definition of small and large classes varies across studies, for the purpose of this study “small” will refer to class size under 30 and “large” will refer to class size over 70.

In higher education, researchers have examined correlations between class size and student achievement (Bandiera, Larcinese, & Rasul, 2010), perceived learning (Cheng, 2011), and satisfaction with the instructor and with the course (Chapman & Ludlow, 2010), with mixed results. The literature review for the current study focused on those studies that explored the relationship between class size and direct assessment of student achievement in higher education. Direct assessment of student achievement through exit tests in math and language indicated that larger classes had a negative effect on student performance in math but not in language learning among first-year students at an Italian university (DePaola, Ponzo, & Scoppa, 2013). The negative effect of the large math class was significantly greater for lower ability students. Direct assessment of student performance on final exams in a small class (fall enrollment, 26; spring enrollment, 25) versus a large class (fall enrollment, 86; spring enrollment, 84) found that students in the small classes scored significantly higher than students in the large classes on final exams in economics (Arias & Walker, 2004). Ho and Kelman (2014) found that assignment to small classes in a law school closed a gender gap in student achievement.

In an analysis of final examination test scores from students in the United Kingdom in classes of differing sizes, Bandiera et al. (2010) found that the effects of class size on student achievement varied on the basis of the difference between class sizes. Achievement in classes of fewer than 20 students was significantly better than achievement in classes with enrollments of 104–211, whereas differences in the middle three quintiles (class sizes between 20 and 104) were negligible. Given Bandiera et al.’s (2010) findings that there were no significant differences in student achievement between students in class sizes that encompassed the two class sizes in our study, we might expect to find no significant differences in our small and large classes. However, instructor observation led us to explore what seemed to be real differences in engagement, achievement, and course completion among students in an introductory biology course (Camfield, McFall, & Land, 2015). In this pilot study, we were interested in trying to answer the questions: Are there differences in student engagement, achievement, and course completion between a small and a large section of introductory biology? If so, what are those differences?

**Method**

We designed this quasi-experimental study with a treatment group of a small section (23 students) and a control group that was a larger section (80 students) of a required introductory biology course for biology, prepharmacy, predental, premed, dental hygiene, sport science, and bioengineering majors. Students enrolled in the treatment or control section using the normal enrollment process, and neither section was among the first to reach its enrollment cap. We controlled as many independent variables as possible, including instructor and teaching methods, assignments and exams, textbook, and time of day of the course. University institutional review board approval was obtained for this study.

**Participants**

Students, typically freshmen, who enrolled in one of two sections of an introductory biology course during the spring semester at a university in the central valley in California were part of the study. Of the 23 students enrolled in the small class, 22 completed the course and received a final grade. Of the 80 students enrolled in the large class, 70 students completed the course and received a final grade. All students were included in the analyses up to the point in which they dropped the course, as
appropriate. In addition to the two sections that were part of this study, there were six other sections of this introductory course taught by other instructors. Students in this course were also assigned to a laboratory section; students enrolled in the two sections of the course under investigation were not necessarily assigned to labs together.

Procedure
To minimize the impact of the time and day of the week the sections were offered, the same instructor taught the two sections on the same days (Monday, Wednesday, Friday) with the smaller class beginning at 11:00 a.m. and ending at 12:15 p.m. and the larger class beginning at 12:30 p.m. and ending at 1:45 p.m. The instructor used the same methods, materials, and exams in both sections. A graduate student observer attended eight class sessions during the semester, helping to confirm that the instruction was similar in both sections. The instructor used the same teaching techniques, such as encouraging class discussion during the lectures and practicing with hypothetical experimental questions and critical thinking activities in both sections of the course, as further described in Camfield et al. (2015).

To assess student engagement, the graduate student researcher observed and recorded student behavior using an interval observation method, commonly used in K–12 education, to estimate individual student and classwide behaviors (Briesch, Hemphill, Volpe, & Daniels, 2015). This method was similar to the Behavioral Observation of Students in Schools (BOSS; Shapiro, 2004) that was designed for observation of individual students. The BOSS has well-documented, high interobserver agreement when used to observe individual students (Shapiro, 2004) and, more recently, classwide behavior (Volpe, Young, Piana, & Zasloffsky, 2012). The graduate student observer had been trained in the BOSS and judged to be highly reliable when observing individual students prior to this study commencing. Shapiro’s definitions of academic engaged behavior and off-task behavior were retained for this study. Starting at the front of the classroom, each student was observed and recorded as displaying on-task/engaged or off-task behavior at the beginning of a 15-second interval. If the student was displaying on-task behavior it was coded as active engagement (i.e., answering a question, writing notes) or passive engagement (i.e., looking at the instructor or lecture materials). If the student was displaying off-task behavior then he or she was observed for the remainder of the 15-second interval, and the student’s off-task behavior was observed and noted as motor (i.e., out of seat), verbal (i.e., talking about nonacademic topics or talking when the instructor was talking), or passive (i.e., looking toward the window, looking away from instructor, notes, or lecture materials). Data were collected systematically on a different student every 15 seconds for a total of 45 minutes per class session (180 possible intervals per observation). For example, a student was not observed a second time until all students had been observed the first time. Observations were conducted for a total of eight class sessions and were collected on the same day for both sections of the course. Observations were not conducted on days when there were exams.

To assess student achievement, student scores on three multipart midterm exams, a fourth multiple-choice exam that covered the material from the third midterm to the last instructional class, a comprehensive multiple-choice final exam, and laboratory exercises were recorded and compared, as were overall final grades in the course. For each multipart midterm, separate scores were recorded for multiple-choice, short-answer, and essay questions. The overall final grade included all exam and lab scores. We measured student completion by the number of students who received a grade other than “incomplete” at the end of the term.

Data analysis
To answer the first and second parts of the research questions, a series of independent sample $t$-tests were conducted. The first set of $t$-tests were conducted on the observation data to examine if there were differences in the percentage of time students were engaged in the class (either active engagement or passive engagement) as well as the percentage of time students were engaged in off-task behaviors (motor, verbal, and passive). There were 180 possible observation intervals per classroom observation and a total of eight classroom observations. A percentage was calculated for each area observed during each observation and then an overall mean percentage was calculated for the small section and the large section. Additionally, a series of independent sample $t$-tests were conducted on the multipart midterm exams, fourth exam, final exam, lab scores, and final grade to discern if the small class differed from the large class. Last, a two-way contingency table analysis, Fisher’s exact test, was used to examine overall achievement and course completion rates between the small and large class.
Results
The first part of the question we asked was: Are there differences in student engagement between a small and a large section of introductory biology? If so, what are those differences? Results presented in Table 1 indicate the mean percentage of behaviors observed in each category as well as the t-test results. Findings based on observation data were significant, indicating that the students in the small class appeared to be more on task and engaged in the class than students in the large class. Students in the small class were observed to be on task an average of 72% of the observations, whereas students in the large class were observed to be on task an average of 61% of the observations. Although students in the small class displayed fewer off-task behaviors overall, including fewer motor and verbal off-task behaviors, they displayed more off-task passive behavior (see Table 1).

The second part of the research question was: Are there differences in student achievement between a small and a large section of introductory biology? If so, what are those differences? The number of points possible for each section of the exams, the mean score, and the t-test results are reported in Table 2. Although the mean of the exams scores appears low on Table 2, these scores are similar to scores in other sections of this course and previous years’ scores. Final grades for the course were derived using a formula to adjust grades, similar to a “curve.” Thus, final grades may appear higher than would be expected given exam scores. Student achievement measures differed between the two sections, and those differences became more pronounced as the semester continued. On the first midterm, only the number of points obtained on the essay was significantly different between the two classes, but by the third midterm the multiple-choice, short-answer, and essay points were significantly different between the two sections. The significant differences were always such that the small class had higher grades than the large class. Students in the small class also had higher lab scores than students in the large class, despite the fact that students in the small class may have been grouped into labs with students in the larger class or with students from sections that did not participate in this study.

The third part of the research question related to overall achievement and course completion. The average of the overall final grades for the small class was a full letter grade higher than that of the large class, which was significant based on t-test results (see Table 2). The average grade point for those who completed the course and earned grades was 2.62 for the small class compared with 1.61 for the large class. The proportion of students who received passing grades (a D+ or higher) compared with nonpassing grades (D or lower) was significant based on the results of a two-way contingency table analysis (Fisher’s exact test, \( p = .01 \)), with no students in the smaller class earning a failing grade and 17 students in the large class earning a failing grade.

The course completion rate also differed between the large and small section of the course. Twenty-two of 23 students (96%) in the small section completed the course and earned a grade, and 70 of 80 students (87.5%) in the large section completed the course and earned a grade. Although a higher proportion of students in the small class completed the course, the results of a two-way contingency table analysis (Fisher’s exact test, \( p = .45 \)) was not significant.

Discussion
The results of this pilot study indicate that reduced class size has the potential to be an effective strategy for improving student engagement, achievement, and course completion in an introductory biology course that has too often been a gatekeeper instead of a gateway for students.

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>Small class</th>
<th></th>
<th>Large class</th>
<th></th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engaged total</td>
<td>71.67</td>
<td>3.115</td>
<td>60.76</td>
<td>6.465</td>
<td>-4.297</td>
<td>14</td>
<td>.001*</td>
</tr>
<tr>
<td>Active engaged</td>
<td>12.43</td>
<td>2.356</td>
<td>8.89</td>
<td>1.970</td>
<td>-3.262</td>
<td>13.574</td>
<td>.006*</td>
</tr>
<tr>
<td>Passive engaged</td>
<td>59.24</td>
<td>3.156</td>
<td>51.88</td>
<td>5.507</td>
<td>-3.280</td>
<td>14</td>
<td>.005*</td>
</tr>
<tr>
<td>Off-task motor</td>
<td>15.97</td>
<td>4.321</td>
<td>29.31</td>
<td>3.892</td>
<td>6.485</td>
<td>13.849</td>
<td>.000*</td>
</tr>
<tr>
<td>Off-task verbal</td>
<td>0.63</td>
<td>0.551</td>
<td>2.57</td>
<td>1.068</td>
<td>4.577</td>
<td>10.475</td>
<td>.001*</td>
</tr>
<tr>
<td>Off-task passive</td>
<td>11.74</td>
<td>2.870</td>
<td>7.36</td>
<td>3.010</td>
<td>-2.975</td>
<td>13.969</td>
<td>.010*</td>
</tr>
</tbody>
</table>

*p ≤ .05.
who intend to be STEM majors. Students in the smaller class were overall more engaged in the class as recorded by an independent observer using a standardized protocol. Although students in the smaller class tended to be off task less often than students in the larger class, the off-task behaviors engaged in were different in each class, with students in the smaller class engaging in less motor and verbal behaviors and more passive behaviors. In a small class it makes sense that students were more likely to appear passively off task as opposed to motorically or verbally off task, as the instructor is closer to the students in a more intimate setting where any motor or verbal off-task behavior would be more obvious and distracting to other students and the instructor.

Not only did students in the smaller class appear more engaged in the class, they also achieved higher grades on exams, with this difference becoming more pronounced as the semester continued. No students in the smaller class received a failing grade in the course compared with 17 students in the larger class. Although students in the landmark study conducted by Seymour and Hewitt (1997) did not mention class size in relation to their decision to switch from a STEM to a non-STEM major, they did indicate that relationships with faculty were important. It is possible that students in the smaller class in our study developed a closer relationship with the instructor, felt supported, and therefore asked questions and sought clarification with less anxiety about exposing their lack of understanding than in a larger class. Although the course completion rate proportions were in the

<table>
<thead>
<tr>
<th>TABLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-test results for test scores, including means and standard deviations in small and large classes.</td>
</tr>
<tr>
<td>Test/measure</td>
</tr>
<tr>
<td>Exam 1 multiple choice (50 points)</td>
</tr>
<tr>
<td>Exam 1 short answer (40 points)</td>
</tr>
<tr>
<td>Exam 1 essay (10 points)</td>
</tr>
<tr>
<td>Exam 2 multiple choice (60 points)</td>
</tr>
<tr>
<td>Exam 2 short answer (30 points)</td>
</tr>
<tr>
<td>Exam 2 essay (10 points)</td>
</tr>
<tr>
<td>Exam 3 multiple choice (60 points)</td>
</tr>
<tr>
<td>Exam 3 short answer (30 points)</td>
</tr>
<tr>
<td>Exam 3 essay (10 points)</td>
</tr>
<tr>
<td>Exam 4 (100 points multiple choice)</td>
</tr>
<tr>
<td>Final (100 points multiple choice)</td>
</tr>
<tr>
<td>Lab score</td>
</tr>
<tr>
<td>Final grade</td>
</tr>
</tbody>
</table>

*p ≤ .05.
expected direction, had the proportion of students in the large class been equal to the proportion of students in the smaller class that passed the course, six more students in the larger class would have passed the class. That equates to six students who do not have to retake the course, consider a change in major, or be affected by other ramifications that result from failing a course.

Limitations
As a pilot study, this was a first step in engaging in research to discover if there were differences in class size as observed by an instructor. As such, there are some factors that may limit the generalizability of this study. As is common in this type of research, it is difficult to control which students enroll in a class at a given time slot. Although we attempted to limit the day and time effect by offering the classes on the same day in the late morning and early afternoon, there may have been reasons (i.e., conflicts with other required or elective classes) that certain students enrolled in one section over the other section. Similarly, a larger sample size would have increased generalizability. It may have been appropriate to use all the biology sections for this course; however, it would have been difficult to control for instructor differences so we opted for the smaller sample size (two sections instead of eight sections) using the same instructor.

The focus of the observation was on student behavior and not on the instructor’s behavior. The graduate student observer noted that the instructor followed the same format in each class and completed the same activities; however, it is possible that we did not note real differences in instruction in the small and large classroom. As Cotner and colleagues noted (Cotner, Loper, Walker, & Brooks, 2013), the design and layout of a classroom can affect the teaching style, even when the instructor is cognizant of teaching the material in the same way in each classroom. Although the layout of the classrooms was similar, the smaller class was in a smaller room and the larger class was in a larger lecture-style classroom. In the larger lecture-style room it is possible that students felt more anonymous or disengaged from the course. This anonymity may have led to their being absent from class or surfing the web or checking social media during class.

Last, a more in-depth analysis of the equivalency of the students in each section would have been beneficial, such as entering GPA, gender, intended major, and class year. Additionally, attendance was not accounted for in this study. Attendance has often been correlated with overall success in a course (Moore et al., 2003).

Implications and future directions
This pilot study suggests that students in a small section of introductory biology are more engaged in the course, achieve at higher rates, and may be more likely to complete the course compared with students in a larger section. The results of this pilot study are already serving as a catalyst for further discussions among the Biology Department faculty. Those discussions are summarized here but described in depth elsewhere (Camfield et al., 2015). Because smaller classes mean more sections and therefore higher costs to the university, it will be important to learn whether some student characteristics are correlated with greater benefits from small classes. For example, do students from some or all of those groups currently underrepresented in STEM fields benefit more from small classes than students from groups who are dominant in STEM fields? Although successful completion of introductory biology courses is itself an important outcome because it means students can continue through their program’s curriculum, it will be useful to examine the future achievement of students in the large and small classes in this study to see whether the benefits of the smaller class extend into future courses. It will also be useful to explore the question: Does better student performance on exams in introductory biology equate to deeper and more lasting learning as measured by better performance in subsequent biology courses? If so, are those benefits different for different groups of students in ways that we can predict and use to enhance student success and completion of STEM programs? The results of this study show promise and raise questions about how best to use small classes to enhance student learning and success.

Given that small classes are not always feasible for the university because of cost, number of faculty within a department, and other factors, it may also be useful to examine instructional strategies that increase student engagement in the classroom so that students are successful in larger classes.

Conclusions
Although class size in and of itself has been studied at various levels of education with differing results, this study shows promise that smaller class sizes in an introductory biology course increase not only academic achievement but also engagement in the course and a greater likelihood of completing the course. The findings indicate that one way we might enhance student success in STEM courses is by limiting class size, but the results also suggest questions for
more research. A single intervention, such as reducing class size, is unlikely to turn a gatekeeper course into a gateway. This project illustrates that a single research project, however, can become a gateway to renewed faculty interest in student learning and achievement, which may ultimately allow students to find introductory courses to be a gateway to success.

Acknowledgments
This work was supported by the University of the Pacific, including funding from the Pacific Fund, The College of the Pacific’s Dean’s College Research Fund, and the Biology Department. We acknowledge Eileen McFall for her contributions to previous versions of this article.

References


**Amy N. Scott** (ascott2@pacific.edu) is an associate professor in Educational and School Psychology. **Delores E. McNair** is an associate professor in Educational and Organizational Leadership. **Jonathan C. Lucas** is a graduate student in Educational and School Psychology, and **Kirkwood M. Land** is an associate professor in the Department of Biological Sciences, all at the University of the Pacific in Stockton, California.