

Article

Increased Preclass Preparation Underlies Student Outcome Improvement in the Flipped Classroom

David Gross,* Evava S. Pietri,^{†‡} Gordon Anderson,[§] Karin Moyano-Camihort,^{||¶} and Mark J. Graham^{†#}

*Department of Biochemistry and Molecular Biology, [§]Department of Computer Science, and ^{||}Center for Teaching and Faculty Development, University of Massachusetts–Amherst, Amherst, MA 01003; [†]Yale Center for Teaching and Learning and [‡]Department of Psychology, Yale University, New Haven, CT 06520; [#]Department of Psychiatry, School of Medicine, Yale University, New Haven, CT 06511

Submitted February 21, 2015; Revised July 20, 2015; Accepted July 28, 2015
Monitoring Editor: Mary Lee Ledbetter

Active-learning environments such as those found in a flipped classroom are known to increase student performance, although how these gains are realized over the course of a semester is less well understood. In an upper-level lecture course designed primarily for biochemistry majors, we examine how students attain improved learning outcomes, as measured by exam scores, when the course is converted to a more active flipped format. The context is a physical chemistry course catering to life science majors in which approximately half of the lecture material is placed online and in-class problem-solving activities are increased, while total class time is reduced. We find that exam performance significantly improves by nearly 12% in the flipped-format course, due in part to students interacting with course material in a more timely and accurate manner. We also find that the positive effects of the flipped class are most pronounced for students with lower grade point averages and for female students.

INTRODUCTION

The flipped college science course provides the majority of standard lecture material online as assigned preclass homework, and thus allows for in-class instruction that is more active and engaging for students (Day and Foley, 2006). An active-learning environment provides benefits that are well-known in science, technology, engineering, and mathematics (STEM) education (President's Council of Advisors on Science and Technology, 2012; Graham *et al.*, 2013;

Freeman *et al.*, 2014), including improved test performance for all students (Haak *et al.*, 2011). For example, a recent meta-analysis of the effect of online instruction blended with in-class instruction suggests that the flipped-class format improves student outcomes by ~13% in STEM classes (Bernard *et al.*, 2014). Jensen *et al.* (2015) have shown that standard and flipped formats of the same class employing active learning do not have significantly different student outcomes, thus pointing toward the active-learning element of the flipped class as the key to improved student outcomes. Other literature also suggests that flipping, or blending, college science classroom instruction benefits student performance, but how this performance gain is attained through students' preparation activities and engagement with course materials is less known (Halverson *et al.*, 2014; Stockwell *et al.*, 2015). In addition, what is known is largely indirect evidence, including self-reported or observational data (Herreid and Schiller, 2012; Eddy and Hogan, 2014; Halverson *et al.*, 2014).

In this paper, we examine what college science students do differently or more intensely in preparing for the more active, flipped-course environment that can be directly linked to better exam performance. The flipped structure prompts students to review material earlier and more often

CBE Life Sci Educ December 1, 2015 14:ar36

DOI:10.1187/cbe.15-02-0040

[†]Present address: Holyoke Community College, Holyoke, MA 01040.

Address correspondence to: David Gross (dgross@biochem.umass.edu).

© 2015 D. Gross *et al.* CBE—Life Sciences Education © 2015 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

than in the standard note-taking, lecture-based course. For example, the structure of the flipped environment may provide students impetus for less crammed, more uniform interaction with the course material throughout the semester. Long-standing cognitive psychology research highlights the benefits of spacing out learning activities over time in contrast to blocked learning, for example, short-term cramming in the lead-up to an exam (Bahrick *et al.*, 1993; Son, 2004). In addition, the weekly preclass assignment, which is necessary to participate in a flipped environment, may provide needed structure to engage with course content more deliberately (Baepler *et al.*, 2014). If so, this increased student persistence in a more timely and accurate manner could account for performance gains in the flipped environment (Preszler *et al.*, 2007; Estrada *et al.*, 2011; Graham *et al.*, 2013).

We hypothesize that the type of preparation structure necessary for the active-learning, flipped-class format is systematically different and leads to better knowledge acquisition and performance when compared with more standard lecture-based formats. For the college science education community, demonstrating how students in science courses attain performance gains in a flipped-classroom environment offers novel and deeper levels of insight into how to structure the college science education experience. Showing that students in a flipped classroom consistently do things differently, earlier, or more often—and in ways that contribute to their higher performance—would make a stronger case that flipping a course is worthwhile and recommended.

Nevertheless, one difficulty in determining how the performance benefits of the flipped classroom are attained is that many STEM education research studies focus on introductory science or distance-education courses (Halverson *et al.*, 2014). While student performance gains in these environments is encouraging, there is inherent variability in beginning college students' interests, motivations, and approaches to studying for and taking exams (Day and Foley, 2006). This makes it challenging to identify precisely what the students are doing differently in their preparation for and interaction with the flipped environment that leads to higher performance. Though there are only a few studies of flipped or blended curriculum innovation that span across introductory and advanced courses or that focus solely on upper-level courses (Preszler *et al.*, 2007), it is in these courses that there is more stability in interest and motivation factors due to students already being committed science majors.

METHODS

Human Subjects Protocol

The Institutional Review Board at the University of Massachusetts–Amherst has approved the work described in this work under protocol number 2013-1714.

Course Structure

The course under study is an upper-level undergraduate one-semester course that covers classical thermodynamics, equilibrium phenomena, reaction kinetics, and statistical and quantum mechanics. The course is required for undergraduate majors in biochemistry and molecular biology, undergraduate BA majors in chemistry, and, for the first 3 yr of the study, for some PhD students in the graduate program in molecular and cellular biology, all at the University of Massachusetts–Amherst. All instances of the course were taught by the same instructor using very similar in-class slide presentations for the standard course and similar activities for the flipped course. Example problems solved in class by the instructor in the standard course were adapted for peer–peer activities in the flipped-format course.

Student outcomes in the course were studied over a period of 5 yr, during which the course was offered in the standard instructor-centered lecture style for the first 3 yr and in the flipped student-centered style for the last 2 yr. All 5 yr had course elements in common. Online homework based in the Online Web Learning (OWL) system (Hart *et al.*, 1999) provided an out-of-class opportunity for students to practice problem solving. An online, interactive textbook, also based in the OWL system (an OWLBook), provided students with background readings, interactive illustrations, and assignable example problems of greater complexity than the online homework. Summative assessments for all years were via three written exams, all of similar format and content. Student populations and class sizes were similar for all 5 yr (Table 1).

The standard-format course met three times per week for 50 min per meeting. Lecture material was presented using a tablet PC that permitted on-screen inking on PowerPoint slides that were projected to the class, and the slide presentation, along with the instructor's verbal comments, were recorded and provided online to students within 24 h of the end of each class. These postlecture recordings, the OWL homework, and the OWLBook, as well as electronic copies of old exams (including solutions) and course handouts, formed the online portion of the standard-format course.

Table 1. Course data for the 5 yr under study^a

AY	Number of students	Number of sections	Number of meetings/week	Number of minutes/week	Active
2007–2008	79	1	3	150	No
2008–2009	88	1	3	150	No
2009–2010	78	1	3	150	No
2011–2012	111	3	1	75	Yes, clickers, peer
2012–2013	133	2	2	100	Yes, TBL

^aAY is the academic year in which the course was offered. Enrollment numbers are for all sections for each AY. The number of meetings per week and the total number of minutes per week spent in class with the instructor are given. The extent and type of active-learning activities in the classroom are shown in the "Active" column. Clickers, use of personal-response hardware in class; peer, student–student interactions facilitated by instructor; TBL, team-based, collaborative student interactions in class.

The flipped-format course met either for one 75-min session per week or for two 50-min sessions per week. This reduced in-class time was supplemented with prerecorded "lectures" available to the students at least a week before class, which increased the online component in the flipped course compared with the standard course. These supplemental lectures were broken into 5- to 20-min chunks on specific topics in the OWLBook. Students were free to view the supplemental lectures or to skip them, as these lectures carried no course credit. The use of less in-class time allowed the instructor to offer more sections of the course, which allowed the class size to remain about constant despite a rapidly increasing total number of students taking the course. All other components of the standard course were present in the flipped course.

Aside from the prerecorded lectures and reduced in-class time for the flipped-format course, a substantial difference between course formats was the increased use of active learning in the flipped classroom. This took the form of peer–peer think–pair–share activities, clicker responses, and example problems for students to work in the once-weekly 75-min sections. In the twice-weekly 50-min sessions, team-based learning (Michaelsen *et al.*, 2004) was used. In this format, teams of five to eight students remained allied throughout the semester. In-class activities included difficult example problems attacked by teams, individual and team readiness assessments on new material, and student explanations of problem solutions on projected whiteboards.

Written exams that were identical in structure and very similar in content form the basis of analysis in this paper. Three exams were used in each iteration of the course, both in the standard and in the flipped format. Exam 1 covered material from the first third of the course and focused mostly on classical thermodynamics. Exam 2 covered material from the middle third of the course and focused on chemical and physical equilibria. Exam 3 covered material from the final third of the course and focused on reaction kinetics and statistical and quantum mechanics. The third exam was not cumulative, and there was no cumulative final exam in the course. The first two exams were given to students over a 90-min time period on an evening during the semester. The third exam was given during the regularly scheduled 120-min final exam period at the conclusion of the semester. Analysis of student access to online material in the weeks before and after each exam was aligned to the day on which each exam was given.

The course exams in this study were all constructed identically, contained assessment questions that probed very similar aspects of course content, and were presented at the same point in the course for each iteration of the course during the 5-yr period being studied and the 2 yr prior to the study. Individual questions on each exam were graded by one person following a grading rubric provided by the instructor, who was the same individual for all iterations of the course. All exams from the three prior years of the course along with the answer keys for each exam were provided to students at the start of each iteration of the course. Given this standardization of exam format and content, we employ the course exams as a measure of student learning outcomes.

Each exam contained five questions. Students were given the option of answering any four or all five of the questions. Most students chose to answer four questions. If four

questions were answered, each one was worth 25 points. If five questions were answered, each one was worth 20 points. Every question was structured with four or five subquestions that, in general, increased in difficulty in progression through the question. Early subquestions typically involved straightforward application of previously described manipulations of data, whereas late subquestions typically involved integration of concepts not explicitly covered in class. Nearly all questions involved mathematical calculations. Students were given formula sheets with the exam and were told not to memorize formulas.

Analyzing Online Material Access

Student usage profiles for online homework (OWL) attempts, correctness, online OWLBook access, and prerecorded lecture video access were examined as a function of time relative to an exam. These data were binned by week, with week 0 including the 7 d before an exam, concluding with the exam date. The 2 wk before (negative values) or following (positive values) the exam week included the seven consecutive days preceding or following the subsequent or prior week. Exams 1 and 2 occurred during the semester at the one-third and two-thirds points of the course, whereas exam 3 occurred during finals week, even though it was not a cumulative exam. Because the third exam terminated the course, the +1- and +2-wk data cover only the first two exams.

Detailed Description of Statistical Analyses

To run the statistical analysis reported in the paper, we used IBM's SPSS software version 19.

Description of Exam Analysis

For our main analysis comparing the flipped with the standard classroom, we ran a mixed-model 2 (class type) \times 3 (exam) analysis of covariance (ANCOVA). We controlled for student grade point average (GPA), which significantly predicted students' exam scores, $F(1440) = 420.39$, $p < 0.001$, partial $\eta^2 = 0.489$. We treated exams 1, 2, and 3 as the within-subjects variables and class type as the between-subjects variable. We found a significant effect for both exam, $F(2439) = 25.41$, $p < 0.001$, partial $\eta^2 = 0.055$, and class type (i.e., standard or flipped), $F(1440) = 59.91$, $p < 0.001$, partial $\eta^2 = 0.12$, but no interaction, $F(2439) = 1.45$, $p < 0.24$, partial $\eta^2 = 0.003$, between the two. Exam scores were higher in the flipped classroom than in the standard classroom. Furthermore, using a Sidak adjustment for multiple comparisons, exam 1 scores were lower than exam 2 scores ($p < 0.001$), and exam 3 scores were lower than both exam 1 ($p < 0.001$) and exam 2 ($p < 0.001$) scores.

Description of Attempts of Online Material Analysis

To analyze attempts at answering online material across the two class types, we ran a mixed-model 2 (class type) \times 5 (week) ANCOVA, controlling for GPA, $F(1445) = 0.60$, $p = 0.44$, partial $\eta^2 = 0.001$. We treated the 3 wk before the exam and the 2 wk after the exam as our within-subjects variable and class type as the between-subjects variable. For the 3 wk before the exam, we averaged across exams 1, 2, and 3, and for the 2 wk after the exam, we averaged across exams 1 and 2. Thus, we looked at students' average attempts to answer the online material 2 wk before, the week of, and 2 wk after an exam.

Description of Accuracy of Online Material Analysis

To analyze accuracy at answering the online questions, we ran a mixed-model ANCOVA, controlling for GPA, $F(1249) = 56.94$, $p < 0.001$, partial $\eta^2 = 0.186$. Again, we treated the 3 wk before the exam and the 2 wk after the exam as our within-subjects variable (averaged across exams) and class type as the between-subjects variable.

Description of Moderation by GPA Analyses

To first examine whether GPA moderated our students' attempts on the online problems in the 3 wk before the exam, we treated GPA as a continuous variable and ran a regression equation predicting attempts from class type (with the flipped class as 1, and the standard class as 0), GPA mean centered, and the interaction between the two.

Description of Gender Analysis

We ran a 2 (class type) \times 2 (gender) \times 3 (exam) mixed-model ANCOVA, again controlling for GPA, $F(1438) = 441.68$, $p < 0.001$, partial $\eta^2 = 0.502$. Although, there was not a significant interaction between class type, gender, and exam, $F(2437) = 1.71$, $p = 0.31$, partial $\eta^2 = 0.005$, we still examined whether gender differences emerged across each class type and the three exams. Across exams 1, 2, and 3 in the standard class, male students performed better than female students (all p values < 0.02). In the flipped class, this difference was no longer significant on exam 1 ($p = 0.088$) and exam 3 ($p = 0.32$).

Student Surveys

To analyze student satisfaction with the flipped-course format, we examined student responses to an anonymous, standardized student evaluation instrument (the University of Massachusetts Student Response to Instruction) that was given to students at the end of each of the standard- and flipped-course instances. We analyzed student responses to questions on 1) their perception of how much they learned and 2) how they ranked the physical chemistry course compared with other courses. Students in the flipped-course instances were offered a separate survey opportunity to probe specific aspects of their satisfaction with the course.

RESULTS

The aim of this study is to investigate how higher exam scores result from different preparation activities of declared science majors in a flipped, upper-level, required biochemistry course titled Elementary Physical Chemistry. The student population for this course is relatively uniform year to year and is composed of individuals who not only have already demonstrated success in a science major but also have shown they are highly motivated to persist in science. Thus, the upper-level course context reduces the potential individual differences in student ability often found in introductory courses, whether for nonmajors or majors. Details of course elements across five student cohorts totaling 464 students (36% female) are described in Table 1. Both the standard and flipped physical chemistry course formats we analyze shared common elements, including online homework (Hart et al., 1999), an online textbook, and three

summative assessment exams that had identical formats and very similar or identical questions between course instances (see the Supplemental Material).

To investigate what students in a flipped-course environment do differently in terms of timing, preparation, and accuracy in answering homework questions, we use a time-series design (i.e., standard to flipped) and a mixed-model ANCOVA to control for potential cohort differences. We compare two versions of the same course, one presented in a standard format that contained several online learning elements but used instructor-based lecturing in class (i.e., a blended course; Bliuc et al., 2007) and the other a flipped course with enhanced in-class activities. We also analyze all results, with students' undergraduate GPA included as an additional control for differences in intrinsic ability. Our measure of student learning outcomes uses three summative exam grades. The averages of each of these exam grades are not statistically significantly different from one another for the three offerings of the standard-course version. The same is true for the averages for the two offerings of the flipped-course version.

Similar to results from other interventions involving a more active and engaged classroom experience (Freeman et al., 2014), when compared with the standard-course format for student cohorts in the flipped-course format we found significant improvements in exam scores for each of the three exams ($p < 0.01$). Figure 1 shows the mixed-model ANCOVA controlling for students' undergraduate GPA. Averaging across all three exams, the flipped class shows an average 11.6% improvement when compared with the standard course ($F(1440) = 59.91$, $p < 0.001$, partial $\eta^2 = 0.12$). This difference occurred for all three exams, and thus there was no interaction between exam or class type ($F(2440) = 1.45$, $p < 0.24$, partial $\eta^2 = 0.003$). However, there was a significant difference between exams ($F(2440) = 25.41$, $p < 0.001$, partial $\eta^2 = 0.055$). Exams 1 and 2 were significantly higher than the third exam ($p < 0.001$), and exam 2 was significantly higher than exam 1 ($p < 0.001$). The third exam covered quantum mechanics, a topic on which students historically do not fare as well. Still, regardless of the individual exam, students did better on exam performance in the flipped classroom than they did in the standard classroom.

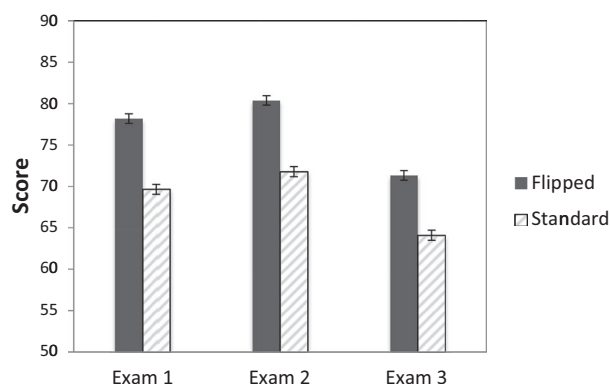


Figure 1. Student exam scores improved significantly in the flipped course as compared with the standard course. We ran a mixed-model ANCOVA while controlling for GPA, and found that students did better in the flipped classroom than the traditional classroom.

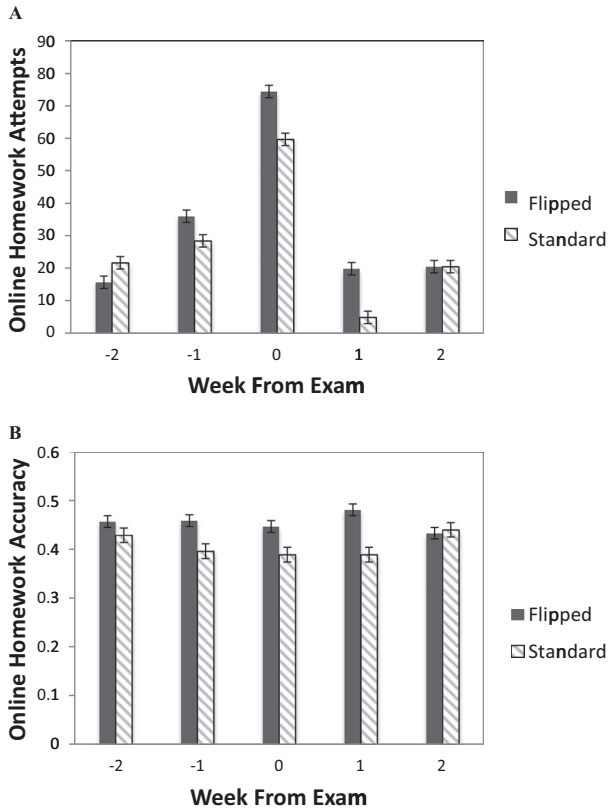


Figure 2. Online homework attempts and accuracy of those attempts as a function of weeks before (negative weeks) or after (positive weeks) an exam. (A) Homework attempts for the two course modes. (B) Accuracy of homework completion for the two course modes.

Exam performance was paired to GPA values binned in six equal size ranges, and these paired data were compared among all course offerings (Supplemental Material Table S1). Comparing years in which the same pedagogy was employed, 28% of the paired scores were significantly different from each other (2-sided *t*-test, $p > 0.05$). In contrast, comparing pairs of years in which one was standard and one was flipped, 43% of the paired exams were statistically significantly different from each other (2-sided *t*-test,

$p > 0.05$), a 50% improvement in student outcomes for the flipped format.

Regarding how this performance increase manifested in the flipped classroom, there are four significant indicators. First, use of the same online material differed substantially between the standard course and the flipped course. Students in the flipped class attempted online homework questions more often, $F(1, 445) = 38.41, p < 0.001$, partial $\eta^2 = 0.08$ (Figure 2A). This difference was not evident for homework completed by the end of the semester (Supplemental Material Table S2). Second, although all students were most likely to attempt homework problems during the week of an exam (p values < 0.001) regardless of class mode, students in the flipped class worked through the online homework more steadily, that is, across the span of weeks, than did students in the standard class. The flipped-class students attempted more homework questions 1 wk before, the week of, and the 2 wk following the exam (p values < 0.001) than did the students in the standard class. Third, students in the flipped class more accurately answered homework problems on average than students in the standard class ($F(1, 249) = 5.84, p < 0.02$, partial $\eta^2 = 0.023$). Fourth, accurately answering homework questions significantly predicted exam scores ($p < 0.001$), indicating that homework accuracy in the 2 wk before the exam mediated the relationship between class type and exam scores ($p < 0.01$). Thus, participating in the flipped class is associated with students more accurately answering homework questions leading up to the exam, which is then significantly related to higher exam scores (Figure 3). Specifically, accuracy partially mediated the relationship between class type and exam scores. We found evidence of mediation both using a traditional Sobel test (Sobel's $z = 3.24, p < 0.01$; Baron and Kenny, 1986) and Hayes' (2013) PROCESS macro with 5000 bootstrap samples to examine the indirect effect. There was a significant indirect effect (i.e., the 95% confidence interval did not cross 0; indirect effect = 1.00, 95% CI = 0.49–1.68). As the figure shows, when we control for accuracy, the relationship between class type and exam scores gets weaker (the β in the parentheses in Figure 3 indicates the effect of class type, controlling for accuracy of OWL problems). However, that relationship is still significant, suggesting that other factors also account for the higher exam scores.

Two individual differences, overall GPA and gender, were also predictive of performance. In the flipped classroom,

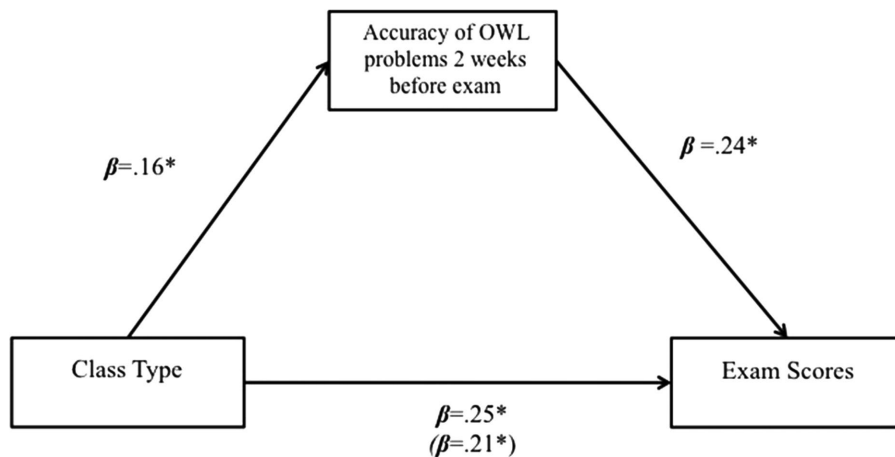


Figure 3. Students in the flipped class answered the online homework problems more accurately in the 2 wk before the exam than students in the standard class. This accuracy in part explains why students in the flipped class had higher exam scores than students in the standard class. Specifically, accuracy partially mediated the relationship between class type and exam scores. β indicates how many SDs the dependent variables (i.e., accuracy, exam score) change per SD increase of the predictor variables (*, $p < 0.05$). β in parentheses indicates the effect of class type, controlling for accuracy of OWL problems.

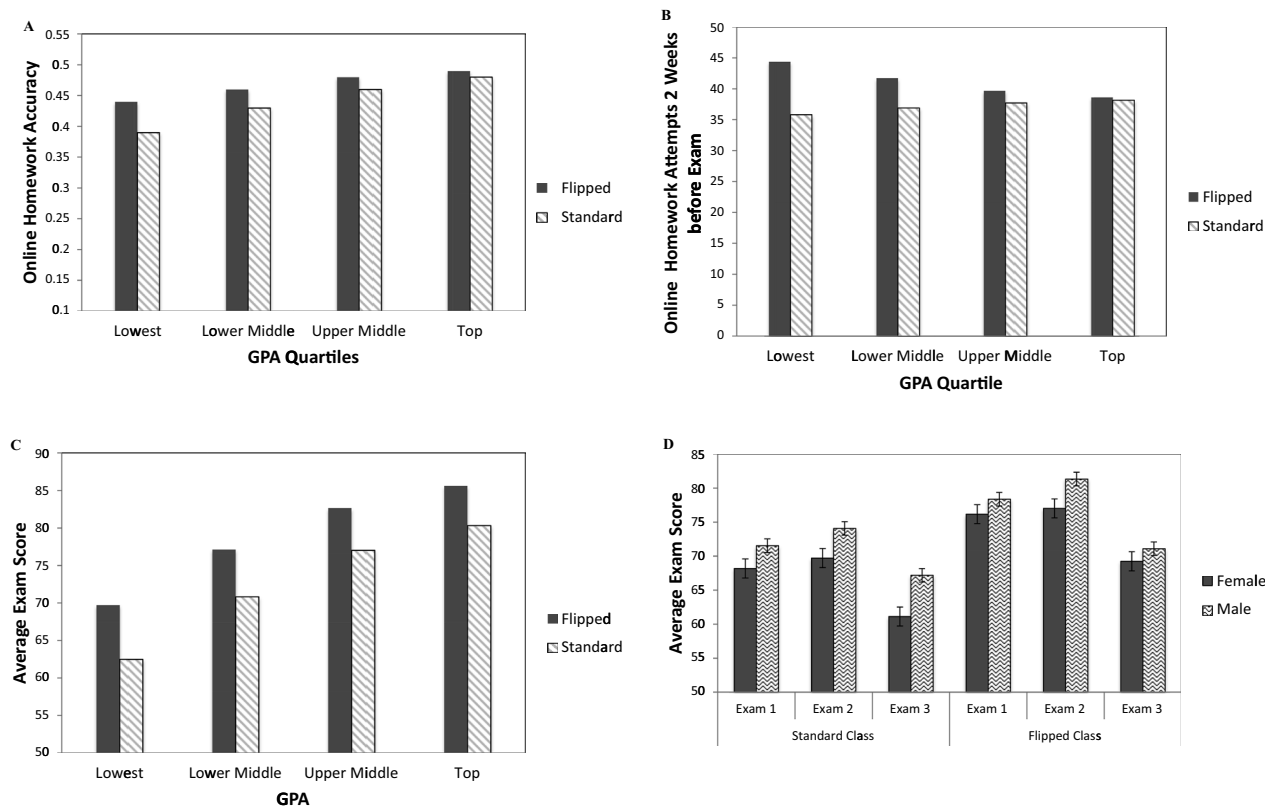


Figure 4. Student GPA correlates with (A) accuracy of completion of online homework 2 wk before an exam, (B) the number of attempts at online homework 2 wk before an exam, and (C) exam score differences between the flipped- and standard-course modes. (D) Female students received significantly lower average exam scores compared with male students in the standard-course format across all three exams, but in the flipped-format course, female and male student scores were not significantly different on exams 1 and 3.

students with lower GPAs showed greater improvement in attempting and accurately answering online homework problems in the weeks immediately before the exam than students with higher GPAs (Figure 4, A and B). Students with lower GPAs also showed greater improvement on exam scores in the flipped- versus standard-course modes (Figure 4C). When examining student attempts and accuracy of answers to online homework problems in the 2 wk leading up to the exam, we saw that students in the lowest GPA quartile in the flipped course showed a larger improvement than students in the highest GPA quartile. This was indicated by a significant interaction between GPA and class type predicting both attempts ($t(444) = -2.68, p < 0.01$) and accuracy ($t(443) = -2.17, p < 0.04$), with students in the lower two GPA quartiles showing the greatest improvement in attempts and accuracy (p values < 0.01).

A gender difference in exam outcomes was also evident. Women demonstrated a greater benefit from the flipped classroom on exam scores than men (Figure 4D). Specifically, there were smaller gender differences on exam scores in the flipped class compared with the standard class. In the standard class across all three exams there was a consistent four to five percentage point difference between female and male students (p values < 0.02). In the flipped class, however, female students did not differ significantly from their male counterparts on exam 1 ($p = 0.09$) and exam 3 ($p = 0.32$). Thus, at least on two out of the three exams, the

flipped-classroom format was alleviating the gender disparity on exam scores.

The flipped-course format relied heavily on online material presented outside of class, including prerecorded lecture videos. Although there was a broad variation in students' access of these videos, there was a significant correlation between average exam score and video access, $r(235) = 0.22, p < 0.001$ (Figure 5A). Students in the flipped-format course with higher exam scores viewed online lecture material more consistently than students with lower exam scores. Students in the top GPA quartile accessed the videos more uniformly across the weeks bordering an exam, in contrast to students in lower GPA quartiles who concentrated their viewing time disproportionately during the week of an exam in a statistically significant manner ($r(235) = 0.24, p < 0.001$; Figure 5B). The relative number of video lecture accesses versus the time in the semester with respect to the three exam dates indicates that the 7 d before an exam are the most likely time for students to access these online materials. There is a significant positive correlation between GPA and video accesses ($r(235) = 0.24, p < 0.001$). Top GPA quartile students opened video lecture files an average of 20.8 times during the semester. Second- through lowest-quartile students opened these files during the semester 13.5, 14.3, and 8.2 times, respectively. (Any set of video accesses by a student that occurred in a time span of < 5 min were counted as a single access for this analysis.)

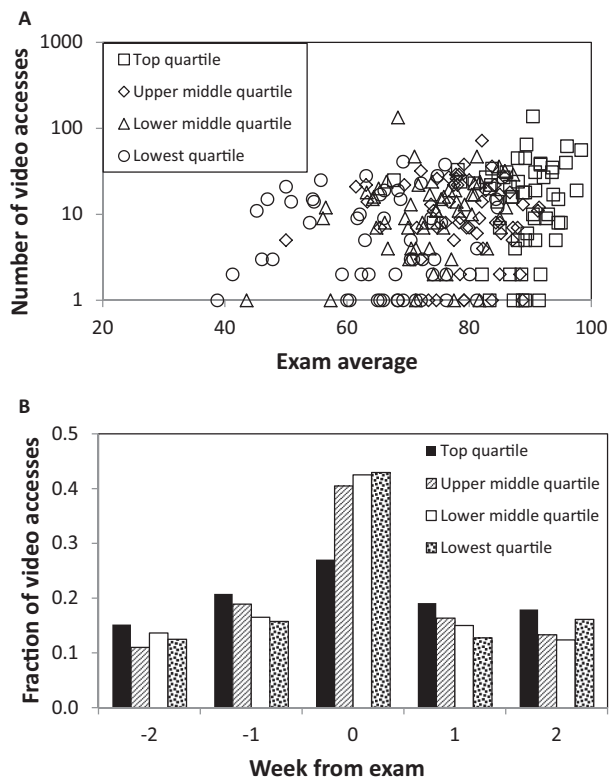


Figure 5. There was substantial correlation between the number of accesses of video recordings and GPA (A) and between the occurrence of those accesses and GPA (B) in the flipped-format course. (A) Note the logarithmic vertical axis. Individual student data are shown. This plot excludes the 14 students who did not access the video lectures. Five of these 14 were GPA fourth-quartile students. The total number of accesses of the online lecture material over the course of the whole semester is plotted vs. the three-exam average for students in the flipped-format course instances.

Student satisfaction with the course was analyzed based on responses to a standard survey instrument answered anonymously at the conclusion of the course. Aggregate responses from students on 1) their perception of how much they learned and 2) how they ranked the course compared with other university courses were compared between standard- and flipped-course instances. Overall, students did not feel that they learned more in the flipped-course version compared with the standard course (despite having better exam scores); however, flipped-format students did rank their course version better overall than did students in the standard-course version (Supplemental Figure S1A).

Students in the two instances of the flipped course were given a second survey that asked their opinions about the flipped-course format (called “blended” in the survey instrument) and the extent to which the course influenced their ability to be independent learners. Supplemental Figure S1B shows that students had a slightly favorable view of the flipped-course format. Students also felt that the flipped-course format had helped them to become more independent learners.

On this same postcourse survey, when asked to select three things to change to improve the course (out of 13 options), 45% of students from the first flipped-course instance

identified “Increase in-class time” as a desired change. Surprisingly, only 19% of the students from the second flipped offering held this view. This difference may relate to the team-based style of instruction in the second offering of the course, though that conclusion requires further study.

DISCUSSION

Our findings demonstrate that there are substantial, positive differences in how students approach a flipped course as compared with a standard-format course. The flipped course encourages students to become more engaged with course material, persist in their learning through more timely and accurate preparation, and, ultimately, perform better. Specifically, this enhanced interaction induces better student preparation for class meetings in the flipped learning environment. More cycles of timely preparation in a flipped class likely improve in-class interactions, which position students to be more accurate in answering online homework problems. This increased accuracy extends to exams, for which grades improve substantially, particularly for lower-GPA students and female students. Because minority groups and women face many external forces impeding their success in the sciences (e.g., anxiety over confirming a stereotype and a threatening environment in STEM; Spencer *et al.*, 1999; Murphy *et al.*, 2007; Moss-Racusin *et al.*, 2012), employing the flipped-class pedagogy in STEM courses is an encouraging pathway for curricular reform efforts aimed at persistence and retention of all students interested in STEM majors. These gains are found despite the fact that face-to-face instructor time with students is 30–50% less for the flipped class.

The initial impetus to convert the course described here from a standard lecture format to the flipped format was to keep class sizes from growing (due to increasing numbers of student majors) without substantially increasing the in-class time commitment of the instructor. This increase in instructor efficiency is counterbalanced by the need for extensive development of online material on the part of the instructor, although that effort rapidly diminishes after the first offerings of the flipped course. The combination of improved student outcomes, at least in the short term, with improved instructor time and classroom use efficiency, makes course flipping an attractive alternative to the standard lecture course.

A key finding from our data is that successful students interact with the online components of a flipped class in a timely manner as compared with students in a standard-format class. That is, the students in the flipped course prepare for class work and avoid the “cramming” style of study for summative assessments, complete the online work more accurately, and perform better on the summative assessments. An important question surrounding this improvement is the role of the flipped-course environment in these improvements. We think that two aspects of the flipped class lead to this improvement: the increase in active student exercises in the classroom coupled to online course content. There is no doubt that active-learning classrooms improve student outcomes, and it has been argued that, for a flipped course, it is active learning that drives improved student outcomes (Jensen *et al.*, 2015; Stockwell *et al.*, 2015). In our case, we believe that the active flipped classroom leads to a student’s

expectation that attending class will require preparation. Additionally, the active classroom, with point-generating activities included in the class sessions, intrinsically encourages students to attend and participate in the activities. Because the flipped classroom offers a clear and reinforcing online experience in the form of recorded “lectures” aligned with online homework, students are encouraged to prepare before class and well before an exam deadline. The interaction of these two elements of the flipped classroom, we believe, provides the underpinning mechanism that accounts for the improvement in student outcomes for the flipped classroom compared with outcomes for the standard classroom.

Our finding that female students in the standard-format class underperform males while in the flipped-format class they perform equally with males points to a fundamental strength of the flipped classroom: students are exposed to a wider variety of learning tools that allow them to better exploit their learning styles. Similarly, lower-performing students receive additional benefit from the flipped-course format as compared with the standard-course format. This effect may also relate to the availability of the wider variety of available learning tools in the flipped-format course. Having several different ways to interact with course content is more likely to resonate with a larger proportion of students than having only one or two modes of interaction with course content, as is typical in a standard course.

ACKNOWLEDGMENTS

The raw data analyzed in this paper can be obtained from the authors. This work was supported in part by the University of Massachusetts. E.S.P. is supported through the Alfred P. Sloan Foundation. M.J.G. is supported through the Howard Hughes Medical Institute’s Professor Grant to Jo Handelsman and by a National Science Foundation grant (NSF-TUES, 1323258).

REFERENCES

- Baepler P, Walker JD, Driessen M (2014). It’s not about seat time: blending, flipping, and efficiency in active learning classrooms. *Comput Educ* 78, 227–236.
- Bahrack HP, Bahrack LE, Bahrack HS, Bahrack PE (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychol Sci* 4, 316–321.
- Baron RM, Kenny DA (1986). The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Personality Soc Psychol* 51, 1173.
- Bernard RM, Borokhovski E, Schmid RF, Tamim RM, Abrami PC (2014). A meta-analysis of blended learning and technology use in higher education: from the general to the applied. *J Comput High Educ* 26, 87–122.
- Bliuc AM, Goodyear P, Ellis RA (2007). Research focus and methodological choices in studies into students’ experiences of blended learning in higher education. *Internet High Educ* 10, 231–244.
- Day JA, Foley JD (2006). Evaluating a Web lecture intervention in a human–computer interaction course. *IEEE T Educ* 49, 420–431.
- Eddy SL, Hogan KA (2014). Getting under the hood: how and for whom does increasing course structure work? *CBE Life Sci Educ* 13, 453–468.
- Estrada M, Woodcock A, Hernandez PR, Schultz P (2011). Toward a model of social influence that explains minority student integration into the scientific community. *J Educ Psychol* 103, 206–222.
- Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP (2014). Active learning increases student performance in science, engineering, and mathematics. *Proc Natl Acad Sci USA* 111, 8410–8415.
- Graham MJ, Frederick J, Byars-Winston A, Hunter A-B, Handelsman J (2013). Increasing persistence of college students in STEM. *Science* 341, 1455–1456.
- Haak DC, HilleRisLambers J, Pitre E, Freeman S (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science* 332, 1213.
- Halverson LR, Graham CR, Spring KJ, Drysdale JS, Henrie CR (2014). A thematic analysis of the most highly cited scholarship in the first decade of blended learning research. *Internet High Educ* 20, 20–34.
- Hart D, Woolf B, Day R, Botch B, Vining W (1999). OWL: an integrated Web-based learning environment. Paper presented at the International Conference on Math/Science Education and Technology (M/SET 99), held in San Antonio, TX, March 1–4, 1999.
- Hayes AF (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, New York: Guilford.
- Herreid CF, Schiller NA (2012). Case studies and the flipped classroom. *J Coll Sci Teach* 42, 62–66.
- Jensen JL, Kummer TA, Godoy PDdM (2015). Improvements from a flipped classroom may simply be the fruits of active learning. *CBE Life Sci Educ* 14, ar5.
- Michaelsen LK, Knight AB, Fink LD (2004). *Team-Based Learning: A Transformative Use of Small Groups*, Sterling, VA: Stylus.
- Moss-Racusin CA, Dovidio JF, Brescoll VL, Graham MJ, Handelsman J (2012). Science faculty’s subtle gender biases favor male students. *Proc Natl Acad Sci USA* 109, 16474–16479.
- Murphy MC, Steele CM, Gross JJ (2007). Signaling threat: how situational cues affect women in math, science, and engineering settings. *Psychol Sci* 18, 879–885.
- President’s Council of Advisors on Science and Technology (2012). *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics*, Washington, DC: U.S. Government Office of Science and Technology.
- Preszler RW, Dawe A, Shuster CB, Shuster M (2007). Assessment of the effects of student response systems on student learning and attitudes over a broad range of biology courses. *CBE Life Sci Educ* 6, 29–41.
- Son LK (2004). Spacing one’s study: evidence for a metacognitive control strategy. *J Exp Psychol Learn* 30, 601–604.
- Spencer SJ, Steele CM, Quinn DM (1999). Stereotype threat and women’s math performance. *J Exp Soc Psychol* 35, 4–28.
- Stockwell BR, Stockwell MS, Cennamo M, Jiang E (2015). Blended learning improves science education. *Cell* 162, 933–936.