International Journal of Teaching and Learning in Higher Education 2019, Volume 31, Number 1, 32-49
http://www.isetl.org/ijtlhe/ ISSN 1812-9129

Differentiated Impact of Flipped Instruction: When Would Flipped Instruction Work or Falter?

Wenliang He, Amanda Holton, Hengrui Gu, Mark Warschauer, and George Farkas
University of California, Irvine

This study assessed the impact of flipped instruction on study effort, exam performance, motivation, and perceived class quality in two sections of an introductory chemistry course. Giving frequent assignments and quizzes provided enough incentive to ensure pre-class study compliance, and flipped instruction did not appreciably increase overall study time. However, technology failures early in the class show an important lesson of what can occur when a teaching modality dependent on technology is used. Unlike in our previous study, flipped students underperformed their control counterparts in the final exam. Differentiated treatment effects were identified, as sophomores and females benefited more from flipped instruction. Similar trends were also observed with student letter grades from a subsequent chemistry course. Flipped instruction did not increase student general motivation. Flipped females, however, exhibited stronger end-of-quarter motivation than flipped males. Flipped students perceived the class to be of lower quality and expressed discontent with in-class technology failures and active learning techniques. We reflect upon the resilience of the traditional lecture format and suggest that new pedagogical methods be implemented at a conservative rate to preserve student learning outcomes in the face of implementation issues.

Flipped instruction is a phenomenon that has rapidly gained momentum since 2008, partly owing to its popularization by two high school teachers (Bergmann & Sams, 2008) and by institutions such as Khan Academy (Bishop & Verleger, 2013). The rise of flipped instruction is a reaction to the discontent with the traditional lectures that have been criticized for perpetuating passive knowledge transfer (Prince, 2004). To encourage productive use of knowledge, a variety of teaching techniques have been invented and are collectively known as active learning techniques, e.g., think-pair-share, peer instruction, in-class demonstration, writing-to-learn, problem-based learning, project-based learning etc. By changing an instructor's role from a "sage on the stage" to a "guide on the side," the goal of active learning is to foster conceptual understanding, analytical skill, creativity, and collaboration. Despite growing evidence showing that active learning works (Michael, 2006; Prince, 2004), many researchers have pointed out that adoption of active learning techniques in practice is hindered by the limited class time to deploy them (Bishop & Verleger, 2013; Moravec, Williams, Aguilar-Roca, & O'Dowd, 2010). Flipped instruction solves this dilemma by offloading the instruction of some new material before class to free up class time for active learning with more practice and problem-solving activities. Unlike hybrid or blended learning, this seat time is not reduced, but simply altered to include more active learning components. This can include completing material that would traditionally be thought of as homework, though generally, the students are still assigned some homework assignments. These post-class homework assignments are often decreased in quantity to account for the work done in class. For these reasons, flipped instruction has attracted immense interests from teachers and researchers alike in recent years.

Treatment Effect on Exam Performance

Although only a handful of empirical studies assessing treatment effect (flipped instruction) on student exam performance were published before 2012, recent years have seen a surge in the number of such studies. Focusing solely on higher education, we have found 35 studies (see Supplemental Material A) that have reported enough information for computing effect sizes as measured by Cohen's d. Among them, eight studies showed negative or null impact, eleven showed small effect ($0 < d < 0.3$), eleven showed moderate to large effect ($0.3 \leq d < 1.0$), and five showed surprisingly large effect ($d \geq 1.0$). For the eight studies showing negative or null impact, all results were statistically non-significant with the largest negative effect size of -0.114. In other words, one in four flipped classrooms was about as effective as traditional classrooms, and three in four of them would outperform their traditional counterparts.

While most empirical studies have focused on measuring overall impact, fewer have examined the potential heterogeneity of the treatment effect. Three studies thus far have reported differentiated treatment effect on performance by question type. With an overall main effect of 0.75, Mason, Shuman, and Cook (2013) reported that their flipped instruction was about twice as effective in improving student performance with design-based problems (ES = 1.19) relative to non-design based ones (ES = 0.58).

Touchton (2015) showed that flipped students performed particularly well in more challenging components of the final applied statistics research paper regarding methodology, diagnostics, and research implications. Quint (2015) found that flipped instruction had a stronger impact on conceptual questions (*ES* = 0.47) as compared to procedural ones (*ES* = -0.10). Thus far, few studies have explored the effects of student demographics. Our prior study (He, Holton, Farkas, & Warschauer, 2016) looked into this issue but did not find any interactions between treatment condition and student demographics. Beyond our study, we have failed to identify other studies investigating this issue.

In large lecture halls, flipped classrooms generally require significant use of technology. During adoption, this can put great stress on the institution's infrastructure. Because implementation issues arise when dramatically changing technologies and course design, the effects of these changes on student learning should be considered. We have not found any published study reporting statistical data collected under these circumstances.

Research Questions

The current study is a follow-up to our previous work. Our prior study showed a small but statistically significant, treatment effect (*ES* = 0.192, p = .008). Student survey responses revealed non-compliance to pre-class study as a major implementation issue that we believe led to the small treatment effect and lack of interaction. The primary goal of this study is to continue our quest to measure overall treatment impact and explore moderation effects. It is of interest to see whether including pre-class for-credit quizzes would provide enough incentive to ensure compliance. Moreover, we are also attentive to students' perceptions of the flipped classroom and to any further implementation issues. Finally, our prior study indicated that flipped instruction caused a shift in student workload from post-class to pre-class without appreciably changing the overall out-of-class time working on course material (study time). This study would check if the result is reproducible.

The instructor had previously taught the course in a flipped format, and the major components of the course were unchanged from the previous implementation (He et al., 2016). However, the change in response system from Iclicker to Learning Catalytics introduced an unexpected opportunity to study the effects of common implementation issues that occur when relying on technology to flip courses.

Hence our current study intends to answer the following research questions for the course:

(1) Did flipped students comply with pre-class study requirement, and did they spend more or less time studying outside the classroom?

(2) Did flipped instruction increase student exam performance and motivation? If so, did students of diverse background benefit equally? Did flipped instruction have sustained impact on students' overall performance in a subsequent course?

(3) Did flipped instruction impact perceived overall class quality?

Method

Course Description

The present study was conducted in Fall 2014 in two sections of a first-year general chemistry course taught by the same instructor at a large public university in the western United States. Previously, the instructor has taught the course seven times in three consecutive years using a traditional lecture format. Flipped instruction was implemented and studied for the first time in Fall 2013. In Fall 2014, a new cohort of 607 students was enrolled into two sections. Both sections met three times a week on Mondays, Wednesdays, and Fridays for ten weeks. The control class was scheduled from 1:00 to 1:50 pm, and the treatment class from 2:00 to 2:50 pm. To avoid students taking alternative sections, class attendance was mandatory and was recorded via Learning Catalytics, a cloud-based learning analytics and assessment system, which accounted for 5% of the final grade.

The control courses were taught in a traditional lecture format. Book reading was recommended, though not "assigned" or tightly correlated with the lectures. No accountability measures were taken to ensure that students did read as recommended. In class the instructor lectured for the full class time. The bulk of the lecture was delivered with PowerPoint slides, with more complex problems being worked out on the document camera. A mixture of definitions, introductory concepts, conceptual discussions, and problem-based discussions was used. While the lectures did occasionally pause for reflections and simple questions were given to the students, time was not set aside to allow them to properly solve or think through a problem on their own. No free work time was given for problem solving. Learning Catalytics was used once per class for a simple knowledge-based question. It was typically given halfway through the class period and was used to control for required attendance in the control section.

For each 50-minute class meeting, the treatment students were required to watch about two online videos before class. The videos created for the previous flipped class were reused. From student feedback, five videos were recreated to increase audio quality, and three long videos were split into short ones. The combined length of the videos remained practically unchanged, totaling 53 videos and 514 minutes with most videos within the range of 5–15 minutes (M = 9.70, SD = 5.01). To ensure compliance, each video was accompanied by an assignment, and each class would begin with a quiz with straightforward questions to test on video material. The assignments focused on questions at the level of remembering and understanding information. Students were expected to spend 60 to 90 minutes per week studying before class. The quizzes accounted for 5% of the total grade.

In the flipped section, a typical meeting was divided into three segments. First, the instructor would briefly review pre-class material and go through the pre-class assignment for 5 to 15 minutes. This portion of the course was still "flipped," given that it included a brief two-minute open-note "quiz" to check for understanding and to increase accountability for watching videos. The quiz questions, much like the assignments, were focused at the level of understanding and remembering. The review itself did not solely repeat factual information but aimed to foster conceptual understanding. The instructor would spend another 10 to 15 minutes with two relatively simple problems. These problems asked the students to understand and apply conceptual and procedural problems. Students worked on the problems in small ad hoc groups (typically 2-4 students) and submitted their answers via Learning Catalytics. Finally, the rest of class time would feature two to three increasingly difficult worksheet problems. These ranged over the full breadth of difficulty and complexity depending on the topic being taught and based on the results of the homework and quiz. Speed and difficulty were adjusted based on class needs. The instructor and teaching assistants would roam over the classroom and offer help whenever needed. Students could submit and change answers at any time, and the results were dynamically displayed to the instructor. The collective responses from the class were shown to the students, and the students were given time to discuss within their groups and change their answers if needed. If the majority of the class faltered, the instructor would either provide more hints or adjourn current activities to address common mistakes. Challenge problems were included, but not discussed, to engage student groups who finished problems before the class was ready to move on. Students were required to complete homework after class, which constituted 10% of the total grade.

For both control and treatment sections, homework was given after class, which constituted 10% of the total grade. The assignments in the treatment course were reduced in volume to approximately 90% to account for the work completed before and in class. Homework was delivered via Mastering Chemistry, which has multiple functionalities but was used in this course primarily for homework. Homework was a mixture of conceptual, definition, and problem solving questions, varying in difficulty from simple one-step questions to complex multi-topic and multi-stepped problems.

Participants

In total, 657 students were initially enrolled in the control (N = 313) and treatment (N = 344) sections. During the first class meeting students were informed of the study and were invited to participate. After excluding students who either dropped the class or did not participate in any exams, the effective sample size was 287 students in the control and 320 in the treatment section, among whom most agreed to participate in the study (i.e., 97.56% or N = 280 and 95.94% or N = 307 respectively). Participants' demographics information was collected from the University's Registrar.

Student demographics were similar between sections, and a detailed comparison is shown in Table 1. Students came from 36 different majors and 12 ethnic groups. For simplicity, majors were regrouped into Biology/Chemistry, STEM (i.e. all STEM majors except for Biology and Chemistry), Non-STEM, and Undeclared. Similarly, ethnicity was regrouped into White, Black/Latino, South Asia, East Asia, and Unstated. High school GPA was collected, since the majority were freshmen who took this course as one of their first college-level courses.

Measures

A number of measures, including exam performance, out-of-class study time, motivation, and perceived class quality, were collected from exams and surveys.

Examinations. Three non-cumulative exams in weeks 3, 6, and 9 and one cumulative final exam in week 11 were administered, accounting for 15%, 20%, 20%, and 25% of the total grade respectively. All exams were similar in form and were administered back to back. To avoid cheating, different forms of the exams were used with isomorphic questions. Raw scores were converted into percentages. Students' letter grades were collected from a subsequent chemistry course, where our course is the first one in the sequence. The letter grades were converted into numeric values in such a way that an A+ corresponds to 13 and an F to 1.

Surveys. Five surveys, a pre-survey and four post-surveys (see Supplemental Material B), were delivered to measure students' study effort, motivation, and perceptions. The pre-survey was given after the first meeting. Each of the identical post-surveys was

Table 1
Descriptive Statistics of Demographics, Pre-Survey Results, and Exam Outcomes by Group

	Control (N = 280)	Treatment (N = 307)		
Measure	M (SD) or Percentage (N)	M (SD) or Percentage (N)	t (p) or x^2 (p)	Cohen's d
SAT Math	604.37 (72.03)	600.19 (76.19)	-0.67 (0.506)	-0.056
High School GPA	2.87 (0.62)	2.78 (0.60)	-1.78 (0.076)	-0.148
Chemistry/Biology	51.97% (145)	63.40% (194)	11.15 (0.011)	
STEM	11.83% (33)	9.48% (29)		
Non-STEM	7.53% (21)	2.94% (9)		
Undeclared	28.67% (80)	24.18% (74)		
Freshman	88.53% (247)	92.81% (284)	3.38 (0.184)	
Sophomore	8.24% (23)	5.56% (17)		
Junior/Senior	3.23% (9)	1.63% (5)		
Male	43.84% (121)	42.81% (131)	0.06 (0.802)	
Female	56.16% (155)	57.19% (175)		
White	11.11% (31)	16.67% (51)	4.28 (0.370)	
Black/Latino	31.54% (88)	28.43% (87)		
South Asia	27.96% (78)	28.76% (88)		
East Asia	26.52% (74)	23.53% (72)		
Unstated	2.87% (8)	2.61% (8)		
Interest	4.21 (0.93)	4.18 (0.96)	-0.28 (0.779)	-0.032
Utility	5.25 (0.84)	5.22 (0.80)	-0.32 (0.750)	-0.037
Importance	4.79 (0.92)	4.77 (0.94)	-0.31 (0.760)	-0.022
Self-efficacy	4.23 (0.87)	4.24 (0.87)	0.13 (0.893)	0.011
Motivation	4.80 (0.61)	4.79 (0.58)	-0.32 (0.749)	-0.017
Pre-class Study Time	5.27 (4.72)	5.35 (4.40)	0.21 (0.834)	0.018
Post-class Study Time	7.44 (5.50)	6.61 (5.94)	-1.61 (0.108)	-0.145
Midterm1	52.69 (17.54)	51.65 (16.86)	-0.73 (0.468)	-0.060
Midterm2	68.85 (15.14)	70.15 (14.85)	1.05 (0.294)	0.087
Midterm3	61.75 (19.23)	61.61 (17.97)	-0.09 (0.926)	-0.008
Final	67.98 (16.28)	64.70 (15.96)	-2.45 (0.014)	-0.204
Post-course Grade	7.01 (2.84)	6.32 (2.92)	-2.49 (0.013)	-0.239

Note. All estimates are standardized beta coefficients. Standard errors are in parentheses.
$^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

administered three days before the corresponding exam to isolate the results from exam performance. To encourage participation, 0.4 extra credits were rewarded for completing each survey, leading up to two extra credits in total. All survey responses were kept separate from the instructor and not processed until after the quarter. Students were advised by a study information sheet that the instructor would receive a list of participants and would not see any results of the survey until after the final grade deadline, and that all results would be reported only in aggregate. Survey items were framed on a 6-point scale with one being the most negatively keyed and six the most positively keyed responses. The survey response rate was higher (over 85%) in the beginning and lower (slightly below 80%) towards the end, averaging 82.64% (*SD* = 4.44%) in the control and 80.91% (*SD* = 3.93%) in the treatment sections.

Our survey motivation items were adapted from the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich, Smith, Garcia, & McKeachie, 1993). Compliant to the expectancy-value theory (Wigfield & Eccles, 2000), items on interest, utility, achievement values, and self-efficacy from MSLQ were used in our study. Three items measured each construct, whose reliability was assessed by Cronbach's alpha. In all surveys, the averaged alpha was over 0.80 for all constructs. A general motivation measure was hence constructed by averaging the twelve items with an average alpha of 0.89 (range: 0.85–0.92) over the surveys.

To measure study effort, the pre-survey asked students to provide numeric estimates of the average number of hours per week they spent studying before and after class for a typical science or mathematics class. Post-surveys asked for estimated average pre- and post-class study time per week during the intervening weeks between the previous exam and the incoming one.

Four post-surveys asked about students' perceived effectiveness of different instructional avenues. Student ratings on lecture quality and class quality were averaged to construct a measure of the overall class quality with a Cronbach's alpha averaging 0.81 (*SD* = 0.03). Post-surveys also included two items asking about the extent flipped students completed all pre-class videos and assignments. Students' narrative comments were collected from the university-wide end-of-quarter optional instructor evaluation.

Note on compliance measures chosen. While video analytics are often suggested as a compliance measure, we opted against using these types of analytics. Students forced to watch videos can allow them to play in the background while not engaging with the material. Additionally, the assignments were written in a manner that allowed students to use the text book or other resources to answer them. It is also expected that many students may work in groups and get help from fellow students to complete the assignments. Because the questions highlighted all important topics in the video, even completing the assignment with the help of a fellow student would ensure a degree of preparedness for class. Students were encouraged to use the modality that best fit their particular preferences and needs.

Results

Preliminary Comparisons

Group equivalence. Descriptive statistics by section are presented in Table 1. Student demographics and pre-survey results suggest reasonable group equivalence on all measures except for high school GPA and majors. Specifically, the flipped students on average had lower GPA by -0.09 points out of 4.00, which is a small effect in size (*ES* = -0.148, *p* = .076). The treatment section, however, had notably 11.43% more Chemistry/Biology majors, and less STEM, undeclared, and non-STEM majors (i.e., 4.59%, 4.49%, and 2.35% respectively); and the chi-squared test showed statistically significant (*p* = .021) difference in majors. In subsequent ordinary least squares (OLS) regression analyses, student demographics were included to address minor group imbalances.

Outcome comparisons. From Table 1, two-sample *t*-tests showed no significant impact of flipped instruction on all three non-cumulative midterms, as the magnitude of the effect sizes was consistently smaller than 0.10 standard deviations. In the cumulative final exam, flipped students on average underperformed their control counterparts by 3.28% (*ES* = -0.204, *p* = .014), which is close to a half-letter grade difference. Furthermore, in the post-chemistry course, the flipped students also underperformed their control counterparts (*ES* = -0.239, *p* = .013).

Compliance and Study Time

(1) Did flipped students comply with pre-class study requirement and did they spend more or less time studying outside the class?

Compliance. To ensure compliance, each class meeting started with a quiz. Flipped students generally did quite well in the quizzes, indicating a high degree of pre-class study compliance. Survey results corroborated this claim. On average, 83.71% (*SD* = 5.13%) of the flipped students indicated that they often finished all the videos before class, among which 36.11% (*SD* = 2.06%) reported to have always finished them. On the contrary, 16.29% (*SD* = 5.13%) claimed that they were often unable to watch all the videos, among which 2.51% (*SD* = 1.79%) claimed that they never watched videos.

Study time. Table 2 shows the self-reported estimates of pre- and post-class study time for each section. Three midterms and one final exam naturally delimited the class into four periods. Flipped students consistently spent more time before class (ten-week average: *ES* = 0.165, *p* = .055) and less time thereafter (*ES* = -0.194, *p* = .024). As a result, the overall out-of-class study time was roughly the same (*ES* = -0.024, *p* = .768). These results confirmed what we had shown in the previous study that flipped instruction did not put extra burden on students, as increase in pre-class study time was offset by decrease in post-class study effort.

Exam Performance and Motivation

(2) Did flipped instruction increase student exam performance and motivation? If so, did students of

Table 2
Self-reported Out-of-class Study Time in Hours by Section

	Week	Control Mean (SD)	Treatment Mean (SD)	*t*-statistic (*p*)	Cohen's *d*
Before-class	Weeks 1-3	4.641 (3.714)	5.298 (3.363)	2.087 (0.037)	0.186
	Weeks 4-6	5.347 (4.078)	5.822 (3.707)	1.326 (0.185)	0.122
	Weeks 7-8	5.241 (4.005)	6.191 (3.915)	2.563 (0.011)	0.240
	Weeks 9-10	6.039 (4.548)	6.86 (4.293)	1.762 (0.079)	0.186
	Weeks 1-10	5.444 (3.834)	6.043 (3.427)	1.927 (0.055)	0.165
After-class	Weeks 1-3	9.67 (5.595)	8.463 (5.378)	-2.482 (0.013)	-0.220
	Weeks 4-6	9.772 (5.63)	8.694 (5.777)	-2.056 (0.040)	-0.189
	Weeks 7-8	9.381 (5.635)	9.032 (5.637)	-0.662 (0.508)	-0.062
	Weeks 9-10	10.29 (6.709)	9.279 (6.263)	-1.477 (0.141)	-0.156
	Weeks 1-10	9.834 (5.472)	8.805 (5.168)	-2.26 (0.024)	-0.194
Out-of-class	Weeks 1-3	12.566 (9.331)	12.124 (8.839)	-0.588 (0.557)	-0.049
	Weeks 4-6	12.688 (9.763)	11.671 (10.212)	-1.233 (0.218)	-0.102
	Weeks 7-8	11.658 (9.656)	11.979 (10.589)	0.385 (0.701)	0.032
	Weeks 9-10	9.896 (11.646)	10.546 (11.373)	0.682 (0.495)	0.057
	Weeks 1-10	11.943 (8.613)	11.73 (8.795)	-0.295 (0.768)	-0.024

Note. All estimates are standardized beta coefficients. Standard errors are in parentheses.
$^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

diverse background benefit equally? Did flipped instruction have sustained impact on student overall performance in a subsequent course?

Exam performance. To account for minor imbalances over GPA and majors, OLS regression was employed, and the results are shown in Table 3. They are explained in brief here and in more detail in the following paragraphs. The first three models used final exam scores as the dependent variable. In our study, the cumulative final exam was valued more than non-cumulative midterms because it revealed the overall long-term impact of flipped instruction. Moreover, 70.36% (*N* = 197) control and 75.89% (*N* = 233) treatment students were enrolled into a subsequent chemistry course in the following quarter. Their letter grades were used as the dependent variable for models 3.4–3.6 in Table 3. In all six models, continuous variables were standardized, and the estimates are hence standardized beta coefficients that can be interpreted as effect sizes.

Model 1.1 is the main effect model that included student demographics and prior motivation as covariates without adding any interaction terms; non-significant terms were not included in the model. High school GPA and majors were statistically significantly associated with the final exam scores, and the treatment effect was somewhat negative (*ES* = -0.107, *p* = .091). Potential interaction effects were studiously explored, and Model 1.2 suggests that females and sophomores benefited from flipped instruction more than males and freshmen. Specifically, while first-year males in the flipped section did significantly worse than their control counterparts (*ES* = -0.276, *p* = .008), first-year females did better than first-year males (*ES* = 0.249, *p* = .055), and sophomores did remarkably better than freshmen (*ES* = 0.545, *p* = .047) in the treatment condition. By implication, it is second-year females who benefited most from flipped instruction. In fact, by changing the reference groups, the OLS model revealed that second-year females in treatment condition outperformed their control counterparts (*ES* = 0.517, *p* = .060). It is worth mentioning that due to the small presence of sophomores (i.e., 6.84% or *N* = 40), statistical significance as indicated by *p* values should be considered together with the size of the effect that signifies practical importance. Model 1.3 included the interaction between treatment and majors. Although none of the terms were statistically significant, the size of the coefficients suggests the possibility that non-Biology/Chemistry majors did worse in the flipped condition than their Biology/Chemistry counterparts.

Model 1.4 is the main effect model with post-course chemistry grade, as the dependent variable, where flipped students on average did worse than control students (*ES* = -0.129, *p* = .034). Post-course grade, defined as the grade students got in the following course, Chemistry 1B, was determined by registrar data. The same treatment-gender interaction of comparable magnitude (*ES* = 0.233, *p* = .057) reappeared in Model 1.5. The treatment-year interaction was not statistically significant (shown in Model 1.6), most likely due to further reduced sample size, as only 20 sophomores and no juniors or seniors enrolled into the subsequent course. The size of the coefficients, however, echoed the same trend revealed by Model 1.2.

Table 3
Effect of Flipped Instruction on Exam Performance with OLS Models

	Final Exam Score			Post-course Grade		
	Model1.1	Model1.2	Model1.3	Model1.4	Model1.5	Model1.6
(Intercept)	0.086	0.189*	0.156+	0.040	0.115	0.134+
	(0.055)	(0.081)	(0.086)	(0.064)	(0.075)	(0.077)
Treatment	-0.107+	-0.276**	-0.207+	-0.129*	-0.269**	-0.301**
	(0.063)	(0.104)	(0.118)	(0.061)	(0.095)	(0.098)
Motivation (pre-survey)	0.066*	0.061+	0.060+			
	(0.033)	(0.033)	(0.033)			
High School GPA	0.688***	0.685***	0.683***	0.835***	0.834***	0.838***
	(0.035)	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)
SATmath	0.140***	0.146***	0.148***	0.093**	0.094**	0.095**
	(0.035)	(0.036)	(0.036)	(0.035)	(0.035)	(0.035)
Female		-0.162+	-0.168+	-0.175**	-0.302**	-0.315***
		(0.094)	(0.094)	(0.063)	(0.091)	(0.093)
Treatment:Female		0.249+	0.246+		0.233+	0.252*
		(0.129)	(0.131)		(0.122)	(0.123)
Sophomore		-0.161	-0.236			-0.180
		(0.196)	(0.205)			(0.209)
Junior/Senior		0.412	0.288			-0.332
		(0.275)	(0.29)			(0.312)
Treatment:Sophomore		0.545*	0.725*			0.323
		(0.274)	(0.300)			(0.288)
Treatment:Junior/Senior		-0.381	-0.049			NA
		(0.394)	(0.466)			NA
STEM	0.130	0.085	0.185	0.192+	0.189+	0.197+
	(0.122)	(0.126)	(0.168)	(0.109)	(0.109)	(0.114)
Non-STEM	-0.348*	-0.460*	-0.242	-0.572**	-0.609**	-0.469+
	(0.159)	(0.194)	(0.242)	(0.217)	(0.217)	(0.255)
Undeclared	-0.092	-0.094	-0.014	0.015	0.02	0.024
	(0.074)	(0.075)	(0.103)	(0.078)	(0.078)	(0.078)
Treatment:STEM			-0.204			
			(0.243)			
Treatment:Non-STEM			-0.586			
			(0.407)			
Treatment:Undeclared			-0.165			
			(0.146)			
Cases	470	469	469	406	406	406
Adj. R-squared	0.541	0.543	0.543	0.649	0.651	0.650
AIC	980.70	980.12	982.84	744.73	743.02	746.68

Note. All estimates are standardized beta coefficients. Standard errors are in parentheses.
$^{+}$ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Motivation. Shown in Table 4, Model 2.1 is the main effect model with motivation measured by the fourth post-survey as the dependent variable; non-significant demographic covariates were not shown. On average, flipped instruction did not change student motivation to any meaningful extent (*ES* = -0.031, *p* = .705). Model 2.2 shows significant treatment-female interaction (*ES* = 0.338, *p* = .047) and marginally significant GPA-SAT interaction (*ES* = 0.084, *p* = .050). However, the treatment-female interaction was not observed in the second (*ES* = 0.012, *p* = .940 from Model 2.3) and third (*ES* = 0.096, *p* = .544 from Model 2.4) post-surveys.

Perception and Implementation Issues

(3) Did flipped instruction impact perceived overall class quality? Were there further implementation issues?

Table 4
Effect of Flipped Instruction on Motivation with OLS Models

	Model2.1	Model2.2	Model2.3	Model2.4
	Motivation4	Motivation4	Motivation2	Motivation3
(Intercept)	0.065 (0.072)	0.158 (0.106)	0.138 (0.099)	0.191^{+} (0.099)
Motivation (pre-survey)	0.548*** (0.043)	0.524*** (0.044)	0.558*** (0.041)	0.530*** (0.041)
Treatment	-0.053 (0.082)	-0.245^{+} (0.134)	-0.147 (0.125)	-0.175 (0.125)
High School GPA	0.101* (0.045)	0.140** (0.050)	0.101* (0.046)	0.166*** (0.045)
Female		-0.187 (0.122)	-0.091 (0.113)	-0.130 (0.113)
SATmath		-0.080^{+} (0.047)	-0.046 (0.043)	0.024 (0.044)
Treatment:Female		0.338* (0.169)	0.012 (0.158)	0.096 (0.158)
GPA:SATmath		0.084^{+} (0.043)	0.088* (0.039)	0.071^{+} (0.040)
STEM	0.161 (0.153)	0.175 (0.164)	-0.198 (0.154)	-0.096 (0.152)
Non-STEM	-0.436* (0.216)	-0.530* (0.220)	-0.096 (0.196)	-0.857*** (0.205)
Undeclared	-0.248* (0.096)	-0.286** (0.099)	-0.111 (0.093)	-0.227* (0.092)
Cases	422	403	411	396
Adj. R-squared	0.320	0.330	0.370	0.391
AIC	1048.80	994.31	966.31	913.98

Note. All estimates are standardized beta coefficients. Standard errors are in parentheses. $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Perception. Regardless of the introductory nature of this course, 51.55% and 38.92% of the students from the combined sample rated this course as "very" and "adequately" challenging, where the two sections differed little. Students' ratings agreed with exam outcomes, where the average raw scores were consistently less than 70% for both sections across exams. Moreover, in all four periods, flipped students rated the class to be of lower quality (*ES* range: -0.245 – -0.357, *p* value range: 0.009–0.0001).

From post-survey responses, we compared flipped students' ratings of the perceived effectiveness of different instructional avenues. Across periods, in-class problem solving was ranked as the most effective means of learning, followed in order by learning before class, online videos, and in-class group discussion. The textbook and in-class lectures were rated as the least and second least effective means, which is not surprising considering that the textbook was not frequently used and lectures often took only a fraction of class time.

Implementation issues. Student comments from the standard campus-wide instructor evaluation provide additional insight. The positive comments echoed the benefits reported in our previous study, including (a) flexibility for learning at one's own pace, (b) availability of online videos for review before exams, (c) better preparation for class meetings, (d) more opportunities for demonstration and problem solving in class, and (e) more instructor-student interaction. Most importantly, we classified students' negative comments to identify weaknesses in our instruction. Two main sources of criticism emerged from the flipped classroom.

First, flipped students expressed strong frustration with the technology failures in class:

- "Once Learning Catalytics stopped working, we started covering some material."
- "I found the whole Learning Catalytics program to be really distracting. I feel like a lot of lecture time was wasted trying to get it running."

In addition, some flipped students criticized the active learning techniques involved, notably group discussion and peer instruction. The frequency of these comments indicates that perhaps a softer style of active learning might be better suited for the student population studied.

- "The instructor can have more examples of problems in class that she solves with the students before letting them solve other problems themselves. It's hard to apply what we don't know to try to answer the questions."
- "Going through more problems together rather than allowing excess time for group discussion might be better because time is wasted and only a few problems are finished in 50 minutes where as more could be fit in. The idea of giving students time together to try a problem is a nice idea but doesn't always execute the way intended."
- "For a student with a very weak background in Chemistry, being asked questions that I don't know the answer to when seeking help only embarrassed me and makes me not want to ask questions."

Discussion

Compliance and Study Time

Giving assignments associated with each video and low-stakes for-credit quizzes with each class effectively reduced pre-class study non-compliance. This finding agreed with reports from other studies (Foertsch, Moses, Strikwerda, & Litzkow, 2002; Mason et al., 2013; Narloch, Garbin, & Turnage, 2006). On the other hand, although only 16.29% students claimed that they often could not watch all the videos, this small fraction still translates into 50 students. In large undergraduate classes, non-compliance would affect a non-negligible number of students, even though the fraction of students affected might be small. Flipped instructors, therefore, should consider monitoring non-compliance closely, particularly when teaching a class comprised primarily of freshmen whose self-discipline and time-management skills are yet to be developed.

With regard to study effort, our current study reproduced what was observed in our prior study (He et al., 2016): flipped instruction caused a shift in study time from post-class to pre-class without appreciably increasing students' overall workload. By implication, flipped students might benefit from spaced learning (Donovan & Radosevich, 1999). Given some students' opposition to the flipped pedagogy, it is advisable that flipped instructors should communicate this result to the students to dispel the concern that pre-class study would impose extra burden on them.

Exam Performance and Motivation

The presence of interaction effect regarding final exam outcome and post-course grade is an important finding. We believe interaction effect would most likely occur when the treatment conditions agree with the characteristics (e.g., motivation, intellectual capacity, and study habits) of a specific subgroup; others with characters departing from this niche group in varying degrees would thus benefit to lesser extents accordingly. In our case, second year females seemed to be the niche group. Treatment females consistently outperformed their control counterparts in both the final exam (ES = 0.249, p = .055) and post-course grade (ES = 0.252, p = .041), and they showed higher end-of-course motivation (ES = 0.338, p = .047). These three related results provide increased support that this sub-group analysis is of practical importance. In addition, females on average seem to spend more time outside the classroom (ES = 0.149, p = .074) than males did, and flipped females relative to control females spent more time before class (ES = 0.319, p = .069) than flipped males did relative to control males. Similarly, second year students did particularly well in the treatment condition. It is conceivable that sophomores were generally less reliant on instructor-initiated instruction and had stronger self-study, self-discipline, and time management skills. They were hence more receptive to flipped instruction and less hurt by implementation issues, as sophomores rated the class to be of higher quality particularly in the third (ES = 0.577, p = .001) and fourth (ES = 0.400, p = .068) post-surveys.

These results support the conjecture that flipped instruction might be more appropriate for students with strong drive, maturity, and skills. Our prior study suggests that without assignments and quizzes, it would take considerable drive, self-discipline, and self-directed learning skills for students to study before class. Although giving assignments and quizzes spurred students to complete pre-class learning assignments, the same set of attributes is still needed to ensure learning quality. Moreover, these attributes are also crucial for students to actively engage during class. When technology goes awry in a flipped classroom, students with these qualities are arguably less vulnerable to suffer the consequences. Sophomores in our study, for example, might be more mentally mature, self-disciplined, active in self-directed learning, and emotionally less resistant to deviance from traditional lectures, which gave them an edge at every corner over the freshmen who were only high school seniors until recently. As implementation issues with the adoption of new technologies are expected, it is important that instructors implement changes slowly to prevent poor outcomes in the flipped classroom. Our data shows that such conservative adoption is markedly more important in freshmen courses.

Student Perception and Implementation Issues

In this study, flipped students rated the class to be of lower quality. We looked at students' comments for indications regarding implementation issues versus their perception of a flipped classroom.

We believe massive technology failures in the flipped classroom were an important reason for the lower ratings for flipped instruction, even as students singled out flipped class components as most effective for their learning. Evidence for this include the difference in comments and ratings between this course and our previous implementation (He et al., 2016), as well as the implementation that occurred after this study (currently sent out for review). Comments in this implementation singled out technology and the in-class room response system as a hindrance to their learning, while these comments were not present in 2013 and 2015 implementations. Both sections in this study used Learning Catalytics instead of IClickers to facilitate peer instruction and real-time feedback. Each student was assigned a unique IP address and connected to the class via a smartphone or tablet. The control students took the class first and had little issue in this regard. In the treatment section, however, some students (random each day) could not get connected because the control class had used up most of the IP addresses. This situation was not fully resolved until the sixth week. By that time, students were already weary of using the technology. While maintaining the use of Learning Catalytics allowed for complete diagnosis and campus wide adoption of appropriate IT standards, the failure of the class response system instilled negative feelings leading to undesirable consequences.

Second, some flipped students voiced criticisms against certain active learning techniques, notably group discussion and peer instruction. Supported by the ideas of constructivism and zone of proximal development, group work is highly valued by educational researchers and has become a key component in many active learning techniques. Our results suggest, however, that having students work in groups might not be as effective as one would expect, as students often ranked group discussion in the bottom of the list of preferred teaching practices, a finding reported by others as well (Enfield, 2013). Some students expressed frustration with their own limited skills for problem solving and regarded group discussion and peer instruction as ineffective use of class time. Some demanded the instructor to elaborate more on complex concepts and demonstrate solving some problems first before diving into group-based problem solving. These echoed the student reflections in the previous study where technology implementation issues did not occur.

These results prompt us to reflect upon the benefits of flipped instruction and the associated active learning techniques as compared to traditional lectures. Although passive lecturing has its shortcomings, it is likely still the most widely used instructional technique regardless of the variety of novel instructional techniques invented over the past decades to supplant it. We believe the resilience of lecturing owes primarily to its simplicity. In comparison, flipped instruction is a promising, but rather complex, instructional technique that entails making multiple decisions on pre-class and in-class components. In a flipped classroom, for example, an instructor needs to consider the number and length of videos, accompanying practice questions, pre-class quizzes, percentage of lectures retained in class, and the number and types of in-class active learning activities. The more decisions to make, the more it is likely that some step might incur an implementation issue. As a result, we highly recommend that instructors new to the flipped pedagogy should choose fewer and simpler technologies to start with. In addition, it is important to note that many active learning techniques frequently require students to work in groups. Staging group activities also entails making multiple decisions, e.g., the difficulty of the problems, group size, group forming tactic (e.g., getting proper group heterogeneity in skills), and time allotment (i.e., enough time for thorough discussion, but not too much to induce boredom and elicit off-topic conversation). While it is possible for instructors to monitor group work closely in small classes, in large classrooms where complete oversight is possible, student could sit out class time pointlessly, and/or unwittingly reinforce each other's biases and have their prior misconceptions strengthened.

Conclusion

Giving assignments associated with each video and for-credit quizzes with each class effectively reduced pre-class study non-compliance. However, non-compliance could still affect a non-negligible number of students, even though the proportion of students affected might be small. Flipped instructors should therefore consider monitoring non-compliance closely, particularly in large introductory undergraduate classes.

Our current study reproduced what was observed in our prior study that flipped instruction did not appreciably increase the overall study time but only caused a shift in workload, which implies that flipped students might benefit from spaced learning. Flipped instructors could communicate this result to students to dispel the concern that flipped instruction exerts an extra burden on them. Moreover, flipped researchers do not need to reduce class meetings to control for increase in required pre-class study time.

While flipped students on average underperformed their control counterparts in the cumulative final exam (*ES* = -0.204, *p* = .014 by two-sample *t*-test and *ES* = -0.107, *p* = .091 by OLS Model 1.1), strong interaction effects existed between treatment condition and gender as well as year level. Females and sophomores benefited more in the flipped section. Similar trends were also observed with student letter grades in a subsequent chemistry course. The differentiated treatment effect lends support to the conjecture that flipped instruction is more appropriate for students with strong drive, maturity, and learning skills.

Flipped instruction did not increase student motivation throughout the course. The same treatment-gender interaction was observed with the final survey, where flipped females showed much stronger motivation (*ES* = 0.338, *p* = .047) compared to flipped males. However, this interaction effect was not shown with previous surveys. Therefore, the interaction effect might be either appearing gradually or due to random statistical noise. We are currently conducting more analysis on motivation to clarify this issue.

Throughout the course, flipped students rated the class to be of lower quality, as they raised complaints about technology failures in class and about the lack of efficiency with in-class group discussion and peer instruction. The variety of issues associated with our flipped classroom prompted us to reflect upon the resilience of traditional lectures, where its simplicity might be its greatest virtue. We caution against overreliance on complex technologies, suggesting simpler implementation may be best. Institutions are advised that it would be advantageous to trouble-shoot technology in advanced classes where students are not likely to be disadvantaged by technology failures. It is suggested that flipped instructors in first-year introductory courses should start with smaller amounts of active learning, building in complexity until reaching a maximum efficacy for the classroom. For example, instead of diving directly into problem solving, some review and elaboration of difficult concepts is necessary as a gentle warm-up. Rather than using open-ended questions with groups of several students, pairs of students working on a clear problem with timely formative feedback are much more tractable. In fact, for the first several lectures, a partially flipped classroom that retains some portions of lectures is highly recommended and will be adopted and studied in future iterations of this course. Surveys can be delivered early in the second week to gauge student attitudes and identify problems. Once students have displayed favorable attitudes towards the flipped pedagogy, instructors could consider gradually adopting a fully flipped classroom, using more complex technologies or teaching techniques in class, and working with increasingly challenging and open-ended problems. For any novel technology or technique employed, the promise to improve teaching is invariably accompanied by challenges. The most effective methods will depend on the instructor, the students, and the institutional climate: special consideration to each must be given.

References

Bergmann, J., & Sams, A. (2008). Remixing chemistry class. *Learning and Leading with Technology*, *36*(4), 24-27.

Bishop, J. L., & Verleger, M. A. (2013). The flipped classroom: A survey of the research. In *ASEE National Conference Proceedings,* Atlanta, GA.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology, 84*(5), 795-805.

Enfield, J. (2013). Looking at the impact of the flipped classroom model of instruction on undergraduate multimedia students at CSUN. *TechTrends*, *57*(6), 14-27.

Foertsch, J., Moses, G., Strikwerda, J., & Litzkow, M. (2002). Reversing the lecture/homework paradigm using eTEACH® web-based streaming video software. *Journal of Engineering Education, 91*(3), 267-274.

He, W., Holton, A.J., Farkas, G., & Warschauer, M., (2016) The effects of flipped instruction on out-of-class study time, exam performance, and student perceptions. *Learning and Instruction*, *45*(1), 61-71.

Mason, G. S., Shuman, T. R., & Cook, K. E. (2013). Comparing the effectiveness of an inverted classroom to a traditional classroom in an upper-division engineering course. *Education, IEEE Transactions on Education, 56*(4), 430-435.

Michael, J. (2006). Where's the evidence that active learning works? *Advances in Physiology Education, 30*(4), 159-167.

Moravec, M., Williams, A., Aguilar-Roca, N., & O'Dowd, D. K. (2010). Learn before lecture: a strategy that improves learning outcomes in a large introductory biology class. *CBE-Life Sciences Education, 9*(4), 473-481.

Narloch, R., Garbin, C. P., and Turnage, K. D. (2006). Benefits of prelecture quizzes. *Teaching Psychology, 33*(2), 109-112.

Pintrich, P. R., Smith, D. A., García, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and psychological measurement*, *53*(3), 801-813.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education, 93*(3), 223-231.

Quint, C. L. (2015). *A study of the efficacy of the flipped classroom model in a university mathematics* *class* (Unpublished doctoral dissertation). Teachers College, Columbia University.

Touchton, M. (2015). Flipping the classroom and student performance in advanced statistics: evidence from a quasi-experiment. *Journal of Political Science Education*, *11*(1), 28-44.

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68-81.

WENLIANG HE is an associate professor in the Department of Educational Technology at Zhejiang Normal University. His research focuses on investigating the effectiveness of novel instructional technologies and teaching techniques for improving STEM higher education with a special interest in flipped instruction and the associated active learning techniques. He is teaching statistics and machine learning courses and is conducting research in learning analytics and educational data mining.

AMANDA HOLTON is an Assistant Professor of Teaching in the Department of Chemistry at the University of California, Irvine where she teaches general chemistry and scientific literacy classes. Her research focuses on practical applications of pedagogical strategies. These topics include active learning, online learning, and flipped classrooms. She is an HHMI National Academies Education Fellow and UCI Lecturer of the Year 2016. Correspondence concerning this article should be addressed to Amanda Holton, Ph.D., Department of Chemistry, University of California, Irvine, CA 92697, USA. E-mail: abrindle@uci.edu - Phone number: 949-385-3490.

HENGRUI GU was previously a graduate student in the Department of Statistics at University of California, Irvine.

MARK WARSCHAUER is Professor of Education and Informatics in the School of Education at University of California, Irvine. Dr. Warschauer is director of the Digital Learning Lab at UC Irvine, where he works on a range of research projects related to digital media in education. In higher education, his team is looking at instructional pratices in STEM lecture courses, the impact of virtual learning on student achievement, the learning processes and outcomes in MOOCs.

GEORGE FARKAS is Professor of Education and Sociology in the School of Education at the University of California, Irvine. His research focuses on educational inequality and how it can be reduced. This includes research on effective instructional strategies across a variety of subjects and groups of learners. He has published more than 100 articles in peer reviewed journals and is a Fellow of the American Educational Research Association. In the 1990s he developed Reading One-to-One, a paraprofessional tutoring program for early elementary school students who have fallen behind in reading. This was the model for President Clinton's America Reads program which assists college work study students to tutor in local school districts. He is Associate Editor of AERA Open, and his research has been supported by the Institute of Education Sciences, USDOE, NSF, NIH, the Spencer Foundation and other organizations.

Acknowledgements

This paper is based in part upon work supported by the National Science Foundation under Grant Number 1535300.

Appendix A

First Author, Year	Course	Grade Level	Number of Cohorts	Treatment (Sample Size)	Control (Sample Size)	Effect Size (Cohen's *d*)
Day, 2006	UI Design	Upper Level	1	28	18	0.69
Moravec, 2010	Biology	Lower Level	2	752	430	1.42
Papadopoulos, 2010	Statics	Unknown	1	43	11	0.20
Stelzer, 2010	Physics	Lower Level	8	750	750	0.20
Deslauriers, 2011a	Physics	Freshman	1	211	171	2.50
Deslauriers, 2011b	Physics	Upper Level	2	62	48	1.14
Pierce, 2012	Therapeutics	Upper Level		71	missing	0.86
Bishop, 2013	Numerical Methods	Sophomore	1	55	63	0
Choi, 2013	Software	Upper Level	1	38	35	0.11
Guerrero, 2013	Mathematics	Unknown	1	15	29	0.20
Lemley, 2013	Thermodynamics	Upper Level	2	15	23	1.02
Mason, 2013	Control Systems	Senior	2	20	20	0.75
McLaughlin, 2013	Pharmaceutics	Professional	2	162	153	-0.13
Morin, 2013	Engineering	Freshman	2	255	237	0
Wilson, 2013	Statistics	Lower Level	2	45	45	0.54
Albert, 2014	Management	Upper Level	2	321	596	0.19
Baepler, 2014	Chemistry	Lower Level	3	375 / 375	350	0.14 & -0.07
Findlay-Thompson, 2014	Introductory	Unknown	1	30	42	0
Fraga, 2014	English	Unknown	1	25	26	0.36
Ghadiri, 2014	Electronics	Unknown	1	78	50 & 75	0.57 & 0.87
Overmyer, 2014	Algebra	Lower Level	1	136	165	0.22
Rais-rohani, 2014	Statics	Unknown	1	53	57	0.17
Street, 2014	Physiology	Professional	2	177	180	0.29
Willis, 2014	Pre-calculus	Lower Level	2	22	22	-0.03
Winquist, 2014	Statistics	Lower Level	11	53	58	0.36
Wong, 2014	Pharmacology	Professional	2	101	103	0.38
Yelamarthi, 2014	Digital Circuits	Lower Level	2	17	24	0.46
Flynn, 2015	Chemistry	Lower Level	2	398	724	0.11
Hung, 2015	English	Lower Level	1	25	24	1.54
Kennedy, 2015	Calculus	Lower Level	1	77	76	-0.11
Quint, 2015	Calculus III	Upper Level	1	39	41	0.19
Quint, 2015	Calculus III	Upper Level	1	35	36	0.51
Schroeder, 2015	Calculus	Lower Level	1	63	49	0.32
Eichler, 2016	Chemistry	Lower Level	1	452	294	-0.07
He, 2016	Chemistry	Lower Level	1	334	343	0.19

Appendix References

Albert, M., & Beatty, B. J. (2014). Flipping the classroom applications to curriculum redesign for an introduction to management course: Impact on grades. *Journal of Education for Business*, *89*(8), 419-424.

Baepler, P., Walker, J. D., & Driessen, M. (2014). It's not about seat time: Blending, flipping, and efficiency in active learning classrooms. *Computers & Education*, *78*, 227-236.

Bishop, J. L. (2013). *A controlled study of the flipped classroom with numerical methods for engineers* (Doctoral dissertation, Utah State University).

Choi, E. M. (2013). Applying inverted classroom to software engineering education. *International Journal of e-Education, e-Business, e-Management and e-Learning*, *3*(2), 121-125.

Day, J. A., & Foley, J. D. (2006). Evaluating a web lecture intervention in a human–computer interaction course. *Education, IEEE Transactions on*, *49*(4), 420-431.

Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *science*, *332*(6031), 862-864.

Deslauriers, L., & Wieman, C. (2011). Learning and retention of quantum concepts with different teaching methods. *Physical Review Special Topics-Physics Education Research*, *7*(1). Retrieved from http://www.cwsei.ubc.ca/SEI_research/files/Physics/Deslauriers-Wieman_QuantumRetention_PRST-PER_2011.pdf

Eichler, J. F., & Peeples, J. (2016). Flipped classroom modules for large enrollment general chemistry courses: a low barrier approach to increase active learning and improve student grades. *Chemistry Education Research and Practice*, *17*(1), 197-208.

Findlay-Thompson, S., & Mombourquette, P. (2014). Evaluation of a flipped classroom in an undergraduate business course. *Business Education & Accreditation*, *6*(1), 63-71.

Flynn, A. B. (2015). Structure and evaluation of flipped chemistry courses: organic & spectroscopy, large and small, first to third year, English and French. *Chemistry Education Research and Practice*, *16*(2), 198-211.

Fraga, L. M., & Harmon, J. (2014). The flipped classroom model of learning in higher education: An investigation of preservice teachers' perspectives and achievement. *Journal of Digital Learning in Teacher Education*, *31*(1), 18-27.

Ghadiri, K. (2014). Developing and implementing effective instructional stratagems in STEM. Retrieved from https://www.edx.org/sites/default/files/upload/developing_and_implementing_effective_instructional_stratagems_in_stem_f_0.pdf

Guerrero, S., Baumgartel, D., & Zobott, M. (2013). The use of screencasting to transform traditional pedagogy in a preservice mathematics content course. *Journal of Computers in Mathematics and Science Teaching*, *32*(2), 173-193.

He, W., Holton, A.J., Farkas, G., Warschauer, M., (2016) The effects of flipped instruction on out-of-class study time, exam performance, and student perceptions. *Learning and Instruction*, *45*(1), 61-71.

Hung, H. T. (2015). Flipping the classroom for English language learners to foster active learning. *Computer Assisted Language Learning*, *28*(1), 81-96.

Kennedy, E., Beaudrie, B., Ernst, D. C., & St. Laurent, R. (2015). Inverted pedagogy in second semester calculus. *PRIMUS*, *25*(9-10), 892-906.

Lemley, E. C., Jassemnejad, B., Judd, E., Ring, B. P., Henderson, A. W., & Armstrong, G. (2013). Implementing a Flipped Classroom in Thermodynamics. Retrieved from https://peer.asee.org/implementing-a-flipped-classroom-in-thermodynamics.pdf

Mason, G. S., Shuman, T. R., & Cook, K. E. (2013). Comparing the effectiveness of an inverted classroom to a traditional classroom in an upper-division engineering course. *IEEE Transactions on Education, 56*(4), 430-435.

McLaughlin, J. E., Griffin, L. M., Esserman, D. A., Davidson, C. A., Glatt, D. M., Roth, M. T., Gharkholonarehe, N. & Mumper, R. J. (2013). Pharmacy student engagement, performance, and perception in a flipped satellite classroom. *American journal of pharmaceutical education*, *77*(9), 1-8.

Moravec, M., Williams, A., Aguilar-Roca, N., & O'Dowd, D. K. (2010). Learn before lecture: a strategy that improves learning outcomes in a large introductory biology class. *CBE-Life Sciences Education, 9*(4), 473-481.

Morin, M. B., Kecskemety, K. M., & Harper, K. A. (2013). The inverted classroom in a first-year engineering course. Retrieved from https://www.asee.org/public/conferences/20/papers/7230/view

Overmyer, G. R. (2014). *The flipped classroom model for college algebra: effects on student achievement* (Doctoral dissertation, Colorado State University).

Papadopoulos, C., & Roman, A. S. (2010). Implementing an inverted classroom model in engineering statics: Initial results. Retrieved from https://peer.asee.org/implementing-an-inverted-classroom-model-in-engineering-statics-initial-results

Pierce, R., & Fox, J. (2012). Vodcasts and active-learning exercises in a "flipped classroom" model of a renal pharmacotherapy module. *American Journal of Pharmaceutical Education*, *76*(10), 1-5.

Quint, C. L. (2015). *A study of the efficacy of the flipped classroom model in a university mathematics class* (Doctoral dissertation, Teachers College, Columbia University).

Rais-Rohani, M. & Walters, A. (2014). Preliminary assessment of the emporium model in a redesigned engineering mechanics course. *Advances in Engineering Education*, *4*(1), 1-19.

Schroeder, L. B., McGivney-Burelle, J., & Xue, F. (2015). To flip or not to flip? An exploratory study comparing student performance in calculus I. *PRIMUS*, *25*(9-10), 876-885.

Stelzer, T., Brookes, D. T., Gladding, G., & Mestre, J. P. (2010). Impact of multimedia learning modules on an introductory course on electricity and magnetism. *American Journal of Physics*, *78*(7), 755-759.

Street, S. E., Gilliland, K. O., McNeil, C., & Royal, K. (2015). The flipped classroom improved medical student performance and satisfaction in a pre-clinical physiology course. *Medical Science Educator*, *25*(1), 35-43.

Willis, J. A. (2014). *The effects of flipping an undergraduate precalculus class* (Doctoral dissertation, Appalachian State University).

Wilson, S. G. (2013). The Flipped Class a Method to Address the Challenges of an Undergraduate Statistics Course. *Teaching of Psychology*, *40*(3), 193-199.

Winquist, J. R., & Carlson, K. A. (2014). Flipped statistics class results: Better performance than lecture over one year later. *Journal of Statistics Education*, *22*(3), 1-10.

Wong, T. H., Ip, E. J., Lopes, I., & Rajagopalan, V. (2014). Pharmacy Students' performance and perceptions in a flipped teaching pilot on cardiac arrhythmias. *American journal of pharmaceutical education*, *78*(10), 1-6.

Yelamarthi, K., & Drake, E. (2015). A flipped first-year digital circuits course for engineering and technology students. *IEEE Transactions on Education*, *58*(3), 179-186.

Appendix B

Pre-survey (for Both Sections)

1. Please rate how frequently did the following situations happen to you.

	Never
I would study course material in advance to prepare for a class.	Very Rarely
I was under-prepared for a class and hence did not get much from it.	Sometimes
I was over-prepared for a class and hence did not get much from it.	Frequently
I had no clue during group discussions and had to sit the time through pointlessly.	Very Frequently
I finished all pre-assigned readings before attending class.	Always

2. For a typical 4-unit science or math course in a ten-week quarter, please estimate the amount of time you usually spend outside the classroom.
   I usually spend ______ hours per week studying in advance to prepare for the class.
   I usually spend ______ hours per week studying after class.

3. How much do you agree with the following statements regarding your motivation?

	Strongly Disagree Strongly Agree
I am very interested in the content area of this course.	1 2 3 4 5 6
Beyond this quarter, contents from this course will still be useful to me.	1 2 3 4 5 6
For me, being good at chemistry is important.	1 2 3 4 5 6
This course is taking more time than what I would like to put into it.	1 2 3 4 5 6
I am confident that I will do well in this course.	1 2 3 4 5 6
I find studying the course material enjoyable.	1 2 3 4 5 6
This course is taking too much time for others things I would prefer to do.	1 2 3 4 5 6
It’s important for me to do well in this course.	1 2 3 4 5 6
I am going to need what I learn from this course in subsequent courses.	1 2 3 4 5 6
Given my current situation, I am confident of getting a good grade.	1 2 3 4 5 6
Compared to other subjects, being good at chemistry is important for me.	1 2 3 4 5 6
The time I am putting into this course is worth my while.	1 2 3 4 5 6
If I am willing, I can get a high grade in this course.	1 2 3 4 5 6

4. How do you rate the effectiveness of the following approaches to learning?

	Highly Ineffective Highly Effective
read the textbooks	1 2 3 4 5 6
attend lectures in class	1 2 3 4 5 6
watch videotaped lectures online	1 2 3 4 5 6
learn from doing homework and assigned problems	1 2 3 4 5 6
learn with other students outside the classroom	1 2 3 4 5 6

Post-survey (for Treatment Section)

Please answer the following questions based on your learning experience during the ***fourth class period*** from the ***third midterm*** to ***the present***.

1. Please rate how frequently did the following things happen to you.

	Never Happened
I could not finish all the pre-assigned videos/readings in time before class.	Very Rarely Happened
I was under-prepared for class meetings and did not get much from it.	Sometimes Happened
I was over-prepared for class meetings and did not get much from it.	Frequently Happened
I had no clue during group discussions and had to sit the time through pointlessly.	Very Frequently Happened
I finished all pre-assigned videos/readings before attending class.	Always Happened

2. In recent weeks, for each 50-minute class meeting, I spent on average ______ ***minutes*** learning course material (e.g. reading textbook or watching videos) in advance ***before attending class***.

3. In recent weeks, I spent on average _______ ***hours per week*** in total, studying course material and doing homework ***after attending class***.

4. Please rate your agreement with the following statements.

	Strongly Disagree Strongly Agree
I am very interested in the content area of this course.	1 2 3 4 5 6
I find studying the course material enjoyable.	1 2 3 4 5 6
Beyond this quarter, contents from this course will still be useful to me.	1 2 3 4 5 6
I am confident that I will do well in this course.	1 2 3 4 5 6
Given my current situation, I am confident of getting a good grade.	1 2 3 4 5 6
If I am willing, I can get a high grade in this course.	1 2 3 4 5 6
This course is taking more time than what I would like to put into it.	1 2 3 4 5 6
The time I am putting into this course is worth my while.	1 2 3 4 5 6
This course is taking too much time for others things I would prefer to be	1 2 3 4 5 6
For me, being good at chemistry is important.	1 2 3 4 5 6
Compared to other subjects, being good at chemistry is important for me.	1 2 3 4 5 6
It’s important for me to do well in this course.	1 2 3 4 5 6
I prefer this inverted class format to a traditional “lecture” format.	1 2 3 4 5 6
I would prefer to take more science classes using this type of class format.	1 2 3 4 5 6

Please rate the overall quality of the following items	Poor Excellent
Read the textbooks	1 2 3 4 5 6
Attend lectures in class	1 2 3 4 5 6
Watch video lectures online	1 2 3 4 5 6
Learning before class	1 2 3 4 5 6
In-class discussion	1 2 3 4 5 6
In-class problem solving	1 2 3 4 5 6
Overall rating of the course.	1 2 3 4 5 6

5. Open-ended questions (optional)

5.1 What are your major complaints about this course?

5.2 How do you recommend us to improve this course?

Post-survey (for Control Section)

This is the final survey of Chem 1A. You will receive 0.4% extra credits in addition to your overall grade for completing this survey. Your responses to the surveys are strictly confidential and will not be analyzed until after all grades are finalized.

1. Based on your recent learning experience from the third midterm to the present, please rate how frequently did the following situations happen to you..

	Never Happened
I attended the alternate section of the class.	Very Rarely Happened
I prepared for the class in advanced.	Sometimes Happened
I was under-prepared for class meetings and did not get much from it.	Frequently Happened
I was over-prepared for class meetings and did not get much from it.	Very Frequently Happened
I did not prepare for the class in advanced.	Always Happened

2. In recent weeks, for each 50-minute class meeting, I spent on average ______ minutes learning course material (e.g. reading textbook or watching videos) in advance before attending class.

3. In recent weeks, I spent on average _______ hours per week in total, studying course material and doing homework after attending class.

4. Please rate your agreement with the following statements.

	Strongly Disagree Strongly Agree
I am very interested in the content area of this course.	1 2 3 4 5 6
I find studying the course material enjoyable.	1 2 3 4 5 6
Beyond this quarter, contents from this course will still be useful to me.	1 2 3 4 5 6
I am confident that I will do well in this course.	1 2 3 4 5 6
Given my current situation, I am confident of getting a good grade.	1 2 3 4 5 6
If I am willing, I can get a high grade in this course.	1 2 3 4 5 6
This course is taking more time than what I would like to put into it.	1 2 3 4 5 6
The time I am putting into this course is worth my while.	1 2 3 4 5 6
This course is taking too much time for others things I would prefer to be	1 2 3 4 5 6
For me, being good at chemistry is important.	1 2 3 4 5 6
Compared to other subjects, being good at chemistry is important for me.	1 2 3 4 5 6
It's important for me to do well in this course.	1 2 3 4 5 6
I am going to need what I learn from this course in subsequent courses	1 2 3 4 5 6

Please rate the overall quality of the following items	Poor Excellent