International Review of Economics Education xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

International Review of Economics Education

journal homepage: www.elsevier.com/locate/iree

Effects of flipping the principles of microeconomics class: Does scheduling matter?

Lauren Calimeris

Department of Economics, St. John Fisher College, 3690 East Avenue, Kearney 416, Rochester, NY, 14618-3597, USA

ARTICLE INFO

JEL classification:
A2
A22

Keywords:
Flipped classroom
Economic education
Inverted classroom
Undergraduate teaching
Student evaluations
Spacing
Scheduling

ABSTRACT

Flipping the principles of microeconomics classroom significantly improves student-learning outcomes compared to traditionally taught lectures; however, it remains unclear as to if the effect differs depending upon the spacing and scheduling of class meetings. This paper investigates if the quantitative effects of flipping on student outcomes differ by scheduling. It further evaluates if student perceptions about flipping vary depending on the spacing of class meetings. This paper shows that students in flipped classes scored significantly higher on final exams compared to those in a traditional setting, and the effect of flipping did not vary with class spacing. Students in the flipped class setting reported significantly more active learning and were significantly more likely to recommend the professor to other students than those in a traditional setting; however, those in the flipped class that met twice a week for 80 min reported significantly less active learning and were significantly less likely to recommend the professor than those in the traditional setting. The results documented in this study suggest that while the effects of flipping on student learning do not vary with class spacing, student satisfaction does differ with spacing. Therefore, when considering the scheduling of a flipped course, instructors need not worry about the impact of flipping on learning outcomes, but they should consider the importance of student satisfaction.

1. Introduction

As colleges and universities face mounting pressure to improve student learning and increase retention rates, instructors and professors find themselves constantly implementing new teaching techniques in an effort to increase active learning, engage students, and improve their retention of material. One of the more popular techniques being adopted is inverting or flipping the classroom. Flipping the classroom entails reversing the order in which activities of a course take place. That is, traditionally, students are first exposed to course material during class, often in a lecture format, and they work on problems and applications outside of the classroom. In a flipped classroom setting, students first receive the lecture material outside of the classroom, typically either through readings or videos, while during class time, they engage in active and hands-on learning (Bergman and Sams, 2012; Hughes, 2012).

As flipping has gained popularity, research has attempted to quantify the benefits of flipping. A scoping review of early studies on flipping from a variety of disciplines found generally positive effects of flipping on student and faculty satisfaction and weak or indirect evidence of the effect on student learning outcomes. Furthermore, the effectiveness of flipping may depend on the methods used, pedagogy involved in the design of the flip, and details of the flip itself (O'Flaherty and Phillips, 2015). Until recently, most research documented gains in student learning outcomes in STEM disciplines (see, for example, Larson and Yamamoto, 2013; Moravec et al., 2010; Pierce and Fox, 2012). These studies did not include control variables and implemented simple difference in

E-mail address: lcalimeris@sjfc.edu.

https://doi.org/10.1016/j.iree.2018.01.002
Received 31 July 2016; Received in revised form 17 August 2017; Accepted 4 January 2018

means *t*-testing, leaving them unable to assign causality to flipping.

Researchers in economics have attempted to expand upon these to document how flipping affects student outcomes in principles classes. Calimeris and Sauer (2015) implement a treatment-control experiment and document that after a negative adjustment period, students in a fully flipped classroom score roughly 0.5–0.6 standard deviations, or two-thirds to an entire letter grade, higher on midterm and final exams. Similarly, Balaban et al. (2016) find that students in a flipped principles of microeconomics classroom in a large class in a university setting scored 0.2–0.7 standard deviations higher on a final exam than those in a traditionally taught class. They also find evidence of increased student effort during the semester, which may be a mechanism through with flipping is effective. This finding corresponds with Green's (2014) finding that the effectiveness of active learning through an extensive, semester-long simulation is dependent upon students' participation and effort.[1] Caviglia-Harris (2016) documents that students in partially- or fully-flipped courses scored significantly higher on a final exam than did those in a traditional course. Olitsky and Cosgrove (2016) use a differences-in-differences approach and find that students in a flipped-blended learning environment learned significantly more over the course of a semester than those in a traditional course.

Therefore, recent literature has established that over the course of a semester, flipping the principles of microeconomics classroom significantly improves end-of-term student learning outcomes. Questions remain however, as to if the effectiveness of flipping differs depending on class spacing, or the frequency and duration of class meetings, as well as class size. Typically, colleges and universities offer students in introductory microeconomics courses the option to take classes that meet either three times per week for roughly 50–55 min or classes that meet twice a week for 75–80 min. Both options have their costs and benefits with respect to flipping. Meeting less frequently for longer periods may enable students to fully complete activities in class, though they have more to accomplish in each class meeting and between classes. Meanwhile, when meeting more frequently for shorter periods, activities may not wrap up in one class period, breaking up the flow of the activity; however, students have less material to review between classes, and revisiting material from unfinished activities may reinforce concepts and the activities in students minds.

The "spacing effect" is defined in the cognitive psychological literature as "the finding that for a given amount of study time, spaced presentations yield substantially better learning than do massed presentations" (Dempster, 1988). That is, learning is improved when material is presented in shorter, more frequent sessions versus longer, more infrequent sessions. The spacing effect, therefore, would suggest flipping to be more effective in the Monday/Wednesday/Friday classes compared to the Tuesday/Thursday classes.

Interestingly, the educational research on class spacing yields no clear consensus. A number of studies generally find that students in compressed courses or courses meeting only once per week generally do not score worse than those in traditional courses (see, for example, Daniel, 2000; Scott and Conrad, 1991). Meanwhile, other studies confirm the spacing effect in education. In particular, in a financial management course, students who enrolled in classes meeting 3 times per week for 50 min had higher pass rates than those enrolled in a class that met once per week (Henebry, 1997). Similar effects were found in college algebra, with students enrolling in one-day-per-week classes scoring significantly lower than those meeting two or three days per week (Gallo and Odu, 2009). More recently, Miyamoto and Coleman (2015) find that students who increase session counts in MOOCs, holding constant total time, are significantly more likely to complete their certification. Carrington and Houston (2010) find that in intermediate accounting courses, there were no significant differences in grades across four different schedules (one, two, and three days-per-week, or an intensive condensed summer course); however, students those who met three days-per-week received significantly more failing grades or dropped out compared to those in the other schedules. The authors surmise that meeting for too short of a period does not enable students to absorb the information or instructors to effectively communicate the material. To this study's knowledge, the effect of flipping with respect to spacing has yet to be investigated.

Additionally, questions remain as to how the effects of flipping may differ with respect to class size. Research has illustrated that increases in class sizes are associated with decreases in student outcomes in college-level economics courses (Arias and Walker, 2004; Diette and Raghav, 2015) and business courses (Monks and Schmidt, 2011). However, there may be such a thing as too small of a class in the flipped classroom setting. That is, since the flipped classroom involves active learning, often in pairs or groups, classes that are too small may not allow for enough groups, for different sized groups, for big enough groups, or for switching groups.

In addition to determining if flipping affects student learning outcomes, researchers have also investigated student attitudes towards the flipped classroom and have generally found students to respond favorably to the flipped classroom format (Calimeris and Sauer, 2015; Lage et al., 2000; Roach, 2014). Student attitudes towards class are important as they may encourage or discourage students from applying themselves, impacting student learning. Furthermore, untenured professors and instructors have an incentive to satisfy students' needs, wants, and desires as course evaluations often factor into tenure or contract renewal decisions. Therefore, in addition to determining if the effects of flipping differ by course schedule, it is also informative to determine if perceptions of and satisfaction with flipping, as measured by course evaluations, differ by schedule.

This study fills a gap in the literature by investigating the effects of flipping by course spacing. First, this paper establishes an effect of flipping on student test scores similar to that in the previously published literature using five semesters of data for seven classes. Next, this paper discerns if there are differences in the effect of flipping on student outcomes by class meeting duration and frequency, as well as class size. Subsequently, this paper determines how flipping and spacing may impact an instructor's course evaluations.

[1] Green also presents evidence in support of a traditional lecture class as he documents that students in hybrid/online classes or in experimental classes which incorporated an "extensive class simulation" scored significantly lower than those in a traditional lecture/discussion course in principles of macroeconomics; however, Green does not investigate the effectiveness of flipping itself, but rather he examines the effectiveness of different components of a flipped class, not a flip as a whole.

	Flipped	Traditional
Three 55-minute classes (MWF)	Fall 2013	Fall 2012, Fall 2013
Two 80-minute classes (TR)	Spring 2014 (2 sections), Fall 2014	Spring 2013

Fig. 1. Schematic of class flipping and scheduling.

This study is organized as follows: Section 2 describes the methods and data, while Section 3 explains the empirical strategy. Section 4 details the results, while Section 5 leads a discussion on the limitations and findings. Section 6 concludes.

2. Methods & data

The data come from the Fall 2012, Spring 2013, Fall 2013, Spring 2014, and Fall 2014 semesters, and the same instructor taught all of the principles of microeconomics classes. The instructor started flipping her principles of microeconomics class during the Fall 2013 semester. Therefore, there are three classes in the traditionally taught format and four classes are in the flipped format in the dataset.

During the Fall 2012 and Fall 2013 semesters, the instructor taught 55-min sections of principles of microeconomics on Mondays, Wednesdays, and Fridays. During the Spring 2013, Spring 2014, and Fall 2014 semesters, the instructor taught 80-min sections on Tuesdays and Thursdays (see Fig. 1). Therefore, there is variability in both day and duration for both flipped and traditionally taught classes. Class sizes ranged from 5 to 39 people, providing variation in class size.

The reader should note that the instructor did not foresee comparing her principles classes over multiple years. Therefore, while a majority of the material is identical from semester to semester, including notes (though specific examples may have been updated to be more relevant and timely) and some worksheets, the instructor naturally added and adapted study aids for the students over time. All classes had access to the same lecture note base and some in-class worksheets.[2] All classes received practice exams and completed online problem sets, although the online platform for the problem sets switched starting in the Fall 2014 semester in an effort to save students money and to address student complaints about limitations of the previous platform. Starting in the Fall 2013 semester, students also had access to instructor-created reading guides to accompany the textbook.

Students in the traditional classes received the lecture notes live in the classroom where they could ask questions and receive immediate feedback. The instructor also may have shown the students a television or movie clip of the lecture material so they could see it “in the real world.” At the end of each chapter of material, students completed worksheets on that material in class either individually, in pairs, or in groups. Students also read the text and completed online problem sets outside of class for each chapter of material.

Students in the flipped classes received the lecture note content via videos that the instructor created using PowerPoint. All videos had the same format, with some text preloaded (such as definitions), and other items animated on the click of a mouse. The instructor recorded her voice over the slides, and annotated the slides using a pen-tablet technology where appropriate. For example, the slides would animate to have a table or the axes of a graph appear, and she would fill in the table or graph by hand using the pen-tablet technology. Videos also contained graphics and clip art to add to the visual appeal. Videos ranged in length from 4 to 21 min.

Students in the flipped classes were to watch the relevant and assigned videos prior to coming to class.[3] The instructor wrote the upcoming videos, their lengths, and due dates on the board so that students could plan their time accordingly. After watching the videos, students completed either an in-class or an online quiz on the videos in an effort to encourage them to take notes and actually watch the videos. During class time, students engaged in active learning activities, games, or experiments pertinent to the material at hand. Once per week, students spent roughly 50 min working on their online problem sets; in the Monday/Wednesday/Friday classes, students engaged in solving online problem sets every Friday, whereas in the Tuesday/Thursday classes, students worked on these problem sets every Thursday. With respect to spacing and scheduling, all students watched the same number of videos throughout the semester and completed the same activities in class; however, students who met two days per week versus three days per week were assigned more videos to watch between classes (e.g., students who met three days per week may have had to watch 2 videos before each of their classes, for a total of 6, whereas those who met two days per week may have had to watch 3 videos before each of their classes).

To determine the effect of flipping on student outcomes, since the final course grade incorporated different course components depending on the semester and class (e.g., flipped classes had additional quizzes and time to complete online problem sets during class), standardized raw exam scores are utilized.[4] While all classes did not have identical exams, the weight of the exams in the overall course grades were the same across all sections,[5] and the format of the exams was similar across classes: the exam structure included fill in the blank definition, true/false, multiple choice, and short answer questions.

[2] The flipped classes had extra worksheets associated with activities specific to flipping that were not available to those in the traditional class, as they did not complete the same active learning activities during class time due to the nature of the flipped versus traditional models of teaching.

[3] Unfortunately, a flaw in the design of the viewing software did not permit the instructor to monitor if students actually watched the videos, as they could watch them on a tablet or phone, but the software was unable to record that the student watched the video.

[4] Raw scores are used for the standardization to exclude extra credit and curves which do not indicate student learning. Standardization is across all courses and years by exam.

[5] For all courses, the midterm exams and final exam comprised 40% and 30% of the final course grade, respectively. In traditional courses, online problem sets were 20% of the final course grade, whereas in the flipped courses, online problem sets were 15% and quizzes 5% of the final course grade.

To determine student perceptions about the courses, student end-of-semester course evaluation data are analyzed. At the end of each semester, prior to the final exam period, students fill out course evaluations that ask for their opinions on the instructor and course. The answers are based on a Likert-type scale with a value of 1 = "Strongly Disagree" to 7 = "Strongly Agree". Particularly relevant to the analysis of flipped versus traditional learning and student opinions regarding the course structure and instructor include the questions "The instructor encouraged active learning", "The instructor promoted an atmosphere conducive to working and learning", "The instructor managed class time effectively", "The assignments/projects/papers helped me to develop a better understanding of course content", "I learned a lot in this course", "I am satisfied with the quality of instruction in this course", and "I would recommend this instructor to another student."

From Fall 2012–Spring 2014, students were allotted 15 min to fill out this survey in paper format during the class meeting time while the instructor was not in the classroom to avoid instructor influence on the responses. Students also had the option to decline to fill out the survey. Students were note required to complete the entire survey, and some students chose to leave sections blank. Finally, students who missed class on that particular day did not have the opportunity to fill out the survey. The day of the survey was not announced prior to giving it, although students were aware of the time frame during which the evaluations were administered. A volunteer from the class returned the completed surveys to the appropriate location so that the instructor could not alter results.

Starting in the Fall 2014 semester, the college switched the format of the evaluation to an online one, though the content of the survey was identical to the paper version. Students are provided a two-week window at the end of the semester to fill out the survey. Students are not required to fill it out, and no extra credit is given to those who do. Therefore, the reader should note that evaluation data likely suffers from selection bias, regardless of the evaluation format.

Student demographic data, including information on SAT/ACT scores, class year, commuter and first generation scholar status, sex, age, ethnicity, parental education, and household income levels were obtained from the school's records. Exam scores and attendance data come from the instructor's records.

3. Empirical strategy

As this study utilizes data from multiple semesters during which there were changes to the course structure and study material components, this study first investigates if the data find a similar effect of flipping on student outcomes as has been established in the literature. It then investigates if the effect of flipping on student outcomes varies depending on the class spacing, as well as on class size.

To assess the effect of flipping on student outcomes, this study uses a multiple regression analysis that allows for a more precise estimation of the effect that flipping has on student learning compared to a simple difference in means test. Since it includes additional controls for factors that affect test scores, the analysis explains more of the variation in test scores.[6] To first establish if there is an overall effect of flipping on student outcomes, this study estimates the following model, which does not include any controls for scheduling, to determine if the data, while not from a treatment/control setting, paint a similar picture as previous literature has found:

$$EXAM_i = \propto + \beta FLIP_i + S_i^{'}\varphi + X_i^{'}\psi + \varepsilon_i \quad (1)$$

where EXAM is the standardized raw score on student *i*'s first or second midterm or final exam for students $i = 1 \ldots n$. Standardization is across all classes and semesters. FLIP is an indicator variable equal to 1 if the student was enrolled in a flipped class format and equal to 0 if the student was enrolled in a traditional class format. S is a vector of academic controls, including indicator variables for if the student received AP credit, took AP economics, class year, commuter status, and first generation scholar status, as well as controls for number of absences and the student's combined math and reading SAT or converted ACT score.[7] X is a vector of demographic characteristics, including sex, age, ethnicity, household income level, and parental education levels.

To determine if the effect of flipping varies by scheduling, that is if class duration and frequency affect the effect of flipping, this study estimates the following model:

$$EXAM_i = \propto + \beta FLIP_i + \gamma TR_i + \theta FLIP_i{*}TR_i + S_i^{'}\varphi + X_i^{'}\psi + \varepsilon_i \quad (2)$$

where TR is an indicator variable equal to 1 if the student was enrolled in a Tuesday/Thursday (80 min) class and equal to 0 if the student was enrolled in a Monday/Wednesday/Friday (55 min) class. $FLIP{*}TR$ is an interaction capturing the effect of flipping a TR class; therefore θ identifies if there is a differential effect of flipping a Tuesday/Thursday class versus a Monday/Wednesday/Friday class. Thus, if θ is positive (negative), it would indicate that flipping a Tuesday/Thursday course has a greater (smaller) effect on student exam scores than flipping a Monday/Wednesday/Friday course. All other variables are defined as above.

[6] Due to the nature of the data and the research question, this study is unable to use a fixed-effects approach. Ideally, the OLS analysis would incorporate clustered standard errors clustered at the class level; however, as there are only 7 classes in the data, this results in too few clusters, leading to over-rejection. Therefore, the primary analysis in this paper uses robust standard errors, and results using clustered standard errors are provided in the Appendix A.

[7] *AP* stands for "Advanced Placement" and indicates if students took college-level courses in high school, something top-ranking high school students are able to do if their school offers AP courses (courses and number of courses vary by school). AP credit refers to if a student took any AP course besides economics (e.g., biology, calculus, statistics, etc.). AP economics indicates if a student took a college-level economics course while still in high school. Depending on the students' score on a cumulative end of year standardized exam for all AP economics students in the nation, colleges may accept the AP course as a transfer credit. *Class year* refers to whether the student is a freshman, sophomore, junior, or senior. While a majority of students take this class as freshmen, students may opt to take this course at any point in their studies. *First generation status* indicates if the student is the first in his or her family to attend college.

Finally, due to the nature of the data, this study is unable to include an interaction of flipping and class size as the only classes considered "small" were taught in a flipped format. Therefore, to determine if scheduling or class size impact exam scores in a flipped setting, this study estimates the following model for those enrolled in the flipped class only:

$$EXAM_i = \propto + \gamma TR_i + \delta smallclass_i + S_i^{'}\varphi + X_i^{'}\psi + \varepsilon_i \quad (3)$$

where all variables are defined as above. *Smallclass* is an indicator variable equal to one if the student was enrolled in a course with fewer than 30 students.

To determine student responses to the flipped class format and if these responses differ by class duration and meeting frequencies, this paper again utilizes a multiple regression analysis. Unfortunately, because the student response data are anonymous, the author is unable to link the responses to the academic and demographic controls from the learning outcome analysis. However, because the format of the evaluations changed, and because the evaluations may differ for flipped versus traditional class or for different days of the week, this study estimates the following model to attempt to control for the scheduling and evaluation format differences:

$$EVAL_{i,j} = \propto + \beta FLIP_i + \gamma TR_i + \theta FLIP_i{*}TR_i + \vartheta online + \varepsilon_i \quad (4)$$

where EVAL is student *i*'s response to question $j = 1 \ldots 7$ of the evaluation questions: "The instructor encouraged active learning", "The instructor promoted an atmosphere conducive to working and learning", "The instructor managed class time effectively", "The assignments/projects/papers helped me to develop a better understanding of course content", "I learned a lot in this course", "I am satisfied with the quality of instruction in this course", and "I would recommend this instructor to another student." All variables are defined as above, and *online* is a dummy variable equal to one if the evaluation was of the online format.

Finally, during the 2013–2014 academic year, students in the flipped classes had identical online assignments and paper course evaluations. Therefore, variation in evaluation data due to selection, evaluation format, or problem set platform should be negligible or nonexistent for this select group. As such, this study also estimates the differences in evaluations for this select group. Because this study cannot control for any additional variables, this difference is a simple difference in means test.

Table 1
Summary statistics and differences between flipped and traditional class.

	Flipped Class			Traditional Class			Difference			
	Mean	Std Err	n	Mean	Std Err	n	Mean	Std Err	t-stat	N total
Standardized Exam 1 Score	0.28	0.144	73	−0.06	0.096	91	0.34	0.144	2.36	164
Standardized Exam 2 Score	0.17	0.144	71	0.01	0.095	92	0.16	0.144	1.12	163
Standardized Final Exam Score	0.32	0.145	73	−0.10	0.096	93	0.42	0.145	2.89	166
Monday, Wednesday, Friday (=1)	0.42	0.073	73	0.74	0.048	93	−0.32	0.073	−4.35	166
Small Class (=1)	0.21	0.042	73	0.00	0.028	93	0.21	0.042	4.87	166
Senior (=1)	0.04	0.027	73	0.02	0.018	93	0.02	0.027	0.73	166
Junior (=1)	0.07	0.031	73	0.02	0.021	93	0.05	0.031	1.50	166
Sophomore (=1)	0.40	0.060	73	0.08	0.040	93	0.32	0.060	5.39	166
Freshman (=1)	0.49	0.065	73	0.88	0.043	93	−0.39	0.065	−6.02	166
Prior Cumulative GPA	3.15	0.169	23	3.04	0.113	28	0.11	0.169	0.63	51
AP Credit (=1)	0.21	0.066	73	0.25	0.044	93	−0.04	0.066	−0.63	166
AP Economics (=1)	0.11	0.046	73	0.09	0.031	93	0.02	0.046	0.51	166
Math & Reading Combined SAT	1100	17.5	71	1067	11.6	91	33	17.5	1.88	162
Math SAT	566	9.6	71	553	6.4	91	13	9.6	1.35	162
Reading SAT	534	10.8	71	514	7.2	91	20	10.8	1.84	162
ACT converted to SAT	1099	31.1	31	1081	21.3	35	18	31.1	0.56	66
SAT & ACT Combined Total	1103	17.4	73	1068	11.5	93	35	17.4	2.03	166
Absences	1.14	0.287	57	1.58	0.177	93	−0.44	0.287	−1.54	150
Commuter (=1)	0.18	0.058	73	0.15	0.038	93	0.03	0.058	0.47	166
First Generation Scholar (=1)	0.26	0.070	73	0.29	0.047	93	−0.03	0.070	−0.43	166
Female (=1)	0.41	0.076	73	0.37	0.051	93	0.05	0.076	0.59	166
Age (years)	18.34	0.116	73	18.16	0.077	93	0.18	0.116	1.56	166
White (=1)	0.86	0.053	73	0.87	0.035	93	−0.01	0.053	−0.15	166
Black (=1)	0.01	0.021	73	0.02	0.014	93	−0.01	0.021	−0.37	166
Other (=1)	0.07	0.037	73	0.05	0.025	93	0.01	0.037	0.39	166
High Income Household (=1)	0.47	0.078	73	0.45	0.052	93	0.01	0.078	0.18	166
Middle Income Household (=1)	0.23	0.069	73	0.28	0.046	93	−0.05	0.069	−0.68	166
Low Income Household (=1)	0.19	0.064	73	0.23	0.043	93	−0.03	0.064	−0.53	166
Mother's Highest Ed: College (=1)	0.58	0.077	73	0.60	0.051	93	−0.03	0.077	−0.35	166
Mother's Highest Ed: High School (=1)	0.26	0.072	73	0.32	0.047	93	−0.06	0.072	−0.87	166
Mother's Highest Ed: Unknown (=1)	0.15	0.047	73	0.06	0.031	93	0.09	0.047	1.82	166
Father's Highest Ed: College (=1)	0.48	0.079	73	0.49	0.052	93	−0.02	0.079	−0.19	166
Father's Highest Ed: High School (=1)	0.37	0.077	73	0.42	0.051	93	−0.05	0.077	−0.64	166
Father's Highest Ed: Unknown (=1)	0.12	0.046	73	0.08	0.031	93	0.05	0.046	1.04	166

4. Results

Table 1 presents the summary statistics for the sample. The primary sample consists of 166 students, 73 in the flipped classes and 93 in the traditional classes.[8] Students in the flipped class scored significantly higher on the first midterm exam and the final exam, scoring 0.34 and 0.42 standard deviations higher than those in the traditional class, respectively. Flipped class students were also more likely to meet on Tuesday/Thursday courses, and they were more likely to be in a small class.

Students in the flipped classes were significantly less likely to be Freshmen and more likely to be Sophomores. Those in the flipped classes scored significantly higher on the combined SAT and SAT/ACT combination, which includes students who took only the ACT with their ACT score converted to the SAT equivalent score. The differences in total SAT scores appears to be driven by the reading portion of the SAT, as there is no significant difference in the math portion of the SAT. These differences in academic abilities highlight the importance of controlling for these characteristics when estimating the impact flipping has on student outcomes. Simply looking at the difference in exam scores in Table 1 suggests that those in the flipped course scored significantly higher than those in the traditional class for the first and final exams; however, these differences could simply be due to differences in innate abilities.

There were no significant differences in the percent of students who were upper classmen, nor the percentages who took AP courses or AP economics. There were also no significant differences in absences between the classes,[9] or with respect to commuter (students who live off campus) or first generation scholar status.

With respect to demographic characteristics, students do not appear to be significantly different from one another, as there are no significant differences with respect to sex, age, ethnicity, household income, and the majority of parental education categories. Over 85% of students in both classes are white. Those in the flipped class were more likely to have mothers reporting an unknown education level.

Generally, while those in the flipped classes appear to be slightly academically stronger with respect to SAT scores, there were minimal significant differences between the groups, and when enrolling, students were unaware if their class would be taught in the flipped or traditional manner, as there is no indicator during registration. It is possible, however, that after the Fall 2013 or Spring 2014 semesters, prospective students may have heard from previous students about the flipping; however, to the instructor's knowledge, students generally appeared unaware of what a flipped course entailed and did not mention having signed up for the course because of flipping.

4.1. *Effects of flipping on student outcomes*

Table 2 presents the OLS results of the effect of flipping on student outcomes for all classes to determine if there is an overall effect of flipping in general as the data are nonexperimental. Both Models I and II include controls for student academic and demographic characteristics, but Model I does not include a control for the number of absences, as absences may be endogenous, while Model II does include a control for absences.[10]

The controls for both models nearly all exhibit the expected signs, though a majority are not significant. Seniors appear to perform significantly better on all exams in Model I, and on the final exam in Model II, compared to Freshmen. This may be because these students are more adapt at studying and balancing requirements of all of their classes at the end of the semester than their youngest peers. Sophomores also score significantly higher on the first exam than their freshmen counterparts, but the significance disappears as the semester progresses.

AP courses do not significantly impact exam scores, but students who scored higher on the SAT/ACT do significantly better, with a 100 point increase in SAT/ACT conversion scores associated with a 0.4–0.5 standard deviation increase in exam scores for both models. First generation scholars, the first in their families to attend college, appear to do significantly better on the first exam, though the significance weans for the second and final exams. This could be because these scholars are highly motivated at the very beginning of the term, but they may not have the time to devote as much time to studying for each individual class in as the semester progresses. Females generally score higher than males, though this effect is only significant for the final exam, perhaps because females tend to have better time management skills than their male counterparts (Trueman and Hartley, 1996).

Absences are associated with significant decreases in exam scores for all exams in Model II. These results suggest that those who miss class are either hurt by not experiencing the material in class, or they may be students who generally do not put forth full effort into their studies or who struggle more with school.

In all specifications, the main variable of interest, flipping, does not have a significant effect on the first two midterm exams. Those in the flipped classes, however, scored 0.216–0.324 standard deviations higher than those in the traditional class on the final exam. This effect translates to increases of roughly one-third to one-half of a letter grade on an exam. The effect could be due to the ability of students in the flipped class to re-watch lecture videos from earlier in the semester, or it could be that active learning activities in the classroom help students to recall information more easily at a later date (Calimeris and Sauer, 2015). This pattern reinforces and confirms the flipping effect identified in the current literature. The magnitude is in line with Balaban et al. (2016)'s findings of increases of 0.2–0.7 standard deviations, though it is slightly smaller than the effect found in Calimeris and Sauer (2015).

[8] There are 49 students between all of the classes who are missing SAT/ACT information, and these students are excluded from the analysis due to the importance of controlling for academic and innate abilities.

[9] During the Spring 2014 semester, attendance in the flipped courses was not taken.

[10] Attendance was not taken during the Spring 2014 semester, when class sizes were small. Therefore, this study is unable to control for absences and class size simultaneously.

Table 2
OLS Analysis of the effect of flipping on student outcomes, all classes.

	Model I			Model II		
	Exam 1	Exam 2	Final	Exam 1	Exam 2	Final
Flip (=1)	0.081	0.070	0.324**	0.057	−0.002	0.216*
	(0.132)	(0.140)	(0.136)	(0.138)	(0.146)	(0.129)
Absences				−0.119***	−0.185***	−0.178***
				(0.032)	(0.047)	(0.032)
Senior (=1)	0.534*	0.699*	1.331***	0.481	0.508	0.693*
	(0.322)	(0.362)	(0.365)	(0.313)	(0.374)	(0.377)
Junior (=1)	0.226	0.052	0.062	0.102	−0.005	0.279
	(0.374)	(0.482)	(0.464)	(0.439)	(0.533)	(0.458)
Sophomore (=1)	0.330**	0.017	−0.099	0.305*	−0.068	−0.001
	(0.160)	(0.192)	(0.218)	(0.174)	(0.210)	(0.226)
AP or College Credit (=1)	−0.010	0.053	0.183	−0.058	−0.017	0.155
	(0.177)	(0.162)	(0.142)	(0.176)	(0.154)	(0.137)
AP Econ Credit (=1)	−0.041	0.227	0.254	−0.059	0.209	0.119
	(0.219)	−0.175	(0.213)	(0.220)	(0.175)	(0.220)
SAT/ACT Convert Score (=1)	0.005***	0.004***	0.004***	0.005***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Commuter (=1)	0.207	0.109	0.200	0.190	0.086	0.133
	(0.152)	(0.188)	(0.168)	(0.169)	(0.194)	(0.172)
First Generation Scholar (=1)	0.439**	0.405*	0.148	0.471*	0.406	0.194
	(0.217)	(0.237)	(0.226)	(0.256)	(0.255)	(0.236)
Female (=1)	0.200	0.096	0.308**	0.207	0.063	0.220*
	(0.123)	(0.132)	(0.124)	(0.136)	(0.139)	(0.130)
Family & Demographic Controls	Y	Y	Y	Y	Y	Y
Constant	−2.856	−1.172	−0.834	−4.919**	−4.193*	−2.150
	(1.976)	(2.391)	(2.166)	(2.125)	(2.420)	(2.182)
Observations	164	163	166	148	149	150
Adjusted R-squared	0.363	0.222	0.295	0.358	0.282	0.346

Robust standard errors in parentheses. *** p < .01, ** p < .05, * p < .1.

This may be because the data are not from a treatment-control experiment, and some preparation materials, experiments, and the homework platform varied from semester to semester, introducing more variability and noise into the data. Furthermore, the exams also differed from semester to semester, so it could be possible that exams may have varied in difficulty.

Therefore, while these data are not from a controlled experiment, they document a similar effect as to that which has been established in the literature. The question remains as to if flipping affects student learning outcomes differentially for classes depending on scheduling and spacing, or the frequency and duration of the class, as well as for differing class sizes. Table 3 presents the OLS results for Eq. (2) to answer these questions. Model III presents a "simple" regression including the effects of flipping, scheduling, and an interaction only, whereas Models IV and V incorporate academic controls (including class year, AP credit, AP economics credit, SAT/ACT scores, and commuter and first generation status) and family and demographic controls (including sex, age,

Table 3
OLS Analysis of the effect of flipping & scheduling on student outcomes, all classes.

	Model III			Model IV			Model V		
	Exam 1	Exam 2	Final	Exam 1	Exam 2	Final	Exam 1	Exam 2	Final
Flip (=1)	0.049	0.154	0.261	0.014	0.180	0.270	0.013	0.147	0.233
	(0.193)	(0.182)	(0.178)	(0.175)	(0.163)	(0.178)	(0.177)	(0.151)	(0.172)
Tues/Thurs (=1)	0.377	0.062	−0.092	0.207	−0.091	−0.145	0.237	−0.043	−0.091
	(0.237)	(0.226)	(0.242)	(0.195)	(0.204)	(0.205)	(0.188)	(0.188)	(0.193)
Flip*Tues/Thurs (=1)	0.302	−0.020	0.324	0.127	−0.264	0.175	0.090	−0.434	−0.034
	(0.306)	(0.316)	(0.322)	(0.291)	(0.338)	(0.337)	(0.301)	(0.354)	(0.337)
Absence							−0.117***	−0.193***	−0.179***
							(0.032)	(0.046)	(0.033)
Academic controls	N	N	N	Y	Y	Y	Y	Y	Y
Family & Personal Demographic Controls	N	N	N	Y	Y	Y	Y	Y	Y
Observations	164	163	166	164	163	166	148	149	150
Adjusted R-squared	0.092	−0.010	0.039	0.365	0.221	0.288	0.362	0.288	0.338
F-stat	1.13	0.49	3.45**	0.21	0.63	2.81*	0.10	0.86	1.43

Robust standard errors in parentheses.*** p < .01, ** p < .05, * p < .10. F-test is for joint significance of flipping variables. Academic controls include class year, AP credit, AP econ credit, SAT & ACT scores, commuter, and first generation status. Family & personal demographic controls include sex, age, ethnicity, household income, and parental education levels.

Table 4
OLS analysis of the effect of flipping & scheduling on student outcomes, parsimonious specifications, all classes.

	P1			P2			P3		
	Exam 1	Exam 2	Final	Exam 1	Exam 2	Final	Exam 1	Exam 2	Final
Flip (=1)	0.035	0.195	0.326**	0.030	0.169	0.302*	0.046	0.150	0.300*
	(0.170)	(0.162)	(0.164)	(0.173)	(0.167)	(0.167)	(0.169)	(0.170)	(0.159)
Tues/Thurs (=1)	0.187	−0.085	−0.189	0.190	−0.097	−0.206	0.210	−0.080	−0.199
	(0.195)	(0.196)	(0.208)	(0.188)	(0.196)	(0.204)	(0.183)	(0.190)	(0.199)
Flip*Tues/Thurs (=1)	0.067	−0.304	0.063	0.076	−0.221	0.153	0.029	−0.221	0.145
	(0.280)	(0.329)	(0.329)	(0.272)	(0.325)	(0.321)	(0.255)	(0.302)	(0.300)
Class Year & SAT/ACT Scores	Y	Y	Y	Y	Y	Y	Y	Y	Y
Personal Demographic Controls	Y	Y	Y	N	N	N	N	N	N
Family Demographic Controls	Y	Y	Y	Y	Y	Y	N	N	N
Observations	164	163	166	164	163	166	164	163	166
Adjusted R-squared	0.369	0.232	0.283	0.359	0.209	0.280	0.371	0.207	0.290
F-test	0.14	0.77	3.38**	0.16	0.52	3.53**	0.12	0.42	3.39**

Robust standard errors in parentheses. *** p < .01, ** p < .05, * p < .1. F-test is for joint significance of flipping variables. All models include controls for sex. Personal Demographic controls include age, ethnicity, and first generation status. Family Demographic Controls include family income and parental education levels.

ethnicity, household income level, and parental education levels). Model V further controls for absences, which may be endogenous as students often choose whether or not be absent from a class; therefore, Model IV in Table 3 is the preferred specification.[11]

The results from the "simple" regression in Model III indicate that none of the variables of interest are individually significant; however, an F-test on the joint significance of the flipping variables (*Flip* and *Flip*Tues/Thurs*) indicates the terms are positively and jointly significant at the 5% level for the final exam. Including academic and family and demographic controls in Model IV, a similar pattern emerges with flipping terms being jointly significant at the 10% level for the final exam only. In Model V, including controls for absences, the joint significance disappears; however, this should be interpreted with caution, as absences may be endogenous as students choose if and when to miss classes. In all models and specifications, the interaction term is not significant. Therefore, the results from Table 3 indicate that the effects of flipping are independent of class spacing.

As a robustness check, given the relatively small sample size and large number of controls, as well as the potential endogeneity of some controls, such as absences and AP courses, this study re-estimates more parsimonious specifications of Eq. (2) in Table 4. All specifications include controls for SAT/ACT scores, class year, and sex and exclude controls for absences, AP credit, AP Economics, and commuter status. The models are presented from least to most parsimonious. Model P2 additionally excludes ethnicity, age, and first generation status. Model P3 further excludes family demographics including family income and parental education level.

For all specifications in Table 4, the general pattern with respect to the effect of flipping on student exam scores is similar to those presented in Tables 2 and 3, though the magnitude of the effect on the final is slightly larger. The results indicate that flipping is individually significant at the 5% level for model P1 and at the 10% level for models P2 and P3 for the final exam only, suggesting that students in the Monday/Wednesday/Friday flipped classes scored roughly 0.300–0.326 standard deviations higher than those in a traditional Monday/Wednesday/Friday class. The flipping/scheduling interaction results are similar to those presented in Table 2. The interaction of flipping and scheduling is not individually significant for any specification. Thus, the results from all specifications from Tables 3 and 4 suggest that spacing does not appear to impact the effectiveness of the flip.[12] An F-test on all flipping variables indicates they are jointly significant at the 5% level for the final exam only in all three models, confirming again that students in the flipped classes scored significantly higher on the final exam versus those in traditional Monday/Wednesday/Friday courses.

Finally, the effectiveness of flipping may vary with class size. To investigate the effect of class size in flipped classes on student test scores, Table 5 presents the results of estimating Eq. (3) in the flipped classes only. Again, because of the potential endogeneity of some academic controls, Table 5 presents a more parsimonious specification of Eq. (3) similar to model P1 in Table 4. Model FP1 includes academic controls for class year and SAT/ACT scores, personal demographic controls for sex, age, and ethnicity, and family demographic controls for family income and parental education levels. Model FP2 includes controls for first generation scholar.

The results in Table 5 confirm the findings from Tables 2–4 that indicate that scheduling does not have an effect on test scores. Furthermore, class size also does not appear to have a significant effect on test scores. These findings suggest that instructors may not need to worry about small class size when deciding whether or not to flip a class; however, the maximum class size in this study is 39 students, therefore the generalizability of this result to large lecture classes remains unknown.

Thus, the results presented in Tables 2–5 indicate that students in a flipped course perform significantly better on the final exam compared to those in a traditional course, and this effect does not appear to vary depending on the scheduling of the flip. That is, whether one flips a class that meets less frequently for longer periods or more frequently for shorter periods does not differentially

[11] For ease of interpretation, only the variables of interest are included in the remaining regression tables. Tables including the full set of results are available upon request.

[12] The results using clustered standard errors provide weak suggestive evidence that spacing may impact the effectiveness of the flip on the first or second midterm only, depending upon the model specification. There is no impact of spacing on final exam scores in the flipped classes, suggesting that any effect disappears by the end of the course, perhaps after students have time to adjust to the class structure; however, given the small number of clusters, these results should be interpreted with caution. Please refer to the Appendix A for results using clustered standard errors.

Table 5
OLS Analysis of scheduling & class size on student outcomes, flipped classes only.

	FP1			FP2		
	Exam 1	Exam 2	Final	Exam 1	Exam 2	Final
Tues/Thurs (=1)	0.205	−0.481	−0.015	0.256	−0.429	0.072
	(0.258)	(0.388)	(0.293)	(0.257)	(0.406)	(0.300)
Small Class (=1)	−0.170	0.079	−0.094	−0.116	0.147	−0.002
	(0.198)	(0.308)	(0.332)	(0.198)	(0.304)	(0.322)
First Generation Scholar (=1)				0.431	0.454	0.732**
				(0.280)	(0.398)	(0.363)
Class Year & SAT/ACT Scores	Y	Y	Y	Y	Y	Y
Family & Personal Demographic Controls	Y	Y	Y	Y	Y	Y
Observations	73	71	73	73	71	73
Adjusted R-squared	0.338	0.127	0.163	0.349	0.133	0.205

Robust standard errors in parentheses. *** p < .01, ** p < .05, * p < .10. Family & demographic controls include sex, age, ethnicity, family income, and parental education controls.

impact student test scores. It may be that both schedules have their costs and benefits which may balance out when it comes to student learning outcomes. The question remains, however, if students view the flipped courses differently, even if there is no bearing on their learning outcomes.

4.2. Flipping and course evaluations

Table 6 presents the results of how flipping may impact student evaluations compared to a traditional class. The findings suggest that those in the flipped Monday/Wednesday/Friday class rated the instructor and course significantly higher in the instructor encouraging active learning than those in a traditional Monday/Wednesday/Friday course, which is refreshing as a flipped classroom focuses on active learning. They were also significantly more likely to recommend the instructor to another student. These findings suggest that students may enjoy this class format more, which may be a potential mechanism behind the flipping effect on student learning outcomes. There are no significant differences with other categories of the course evaluations.

Students who enrolled in traditional Tuesday/Thursday courses were significantly more likely to respond more favorably to all outcomes except if they felt they learned a lot in the course compared to those in the traditional Monday/Wednesday/Friday courses. One should note that students elect which days of the week to enroll in courses, so these results may be due to a selection issue of the types of students who choose to classes on Tuesdays/Thursdays as opposed to Mondays/Wednesdays/Fridays.

Interestingly, students who were in the flipped classes on Tuesdays/Thursdays were less satisfied with the flipped structure of the class compared to the Monday/Wednesday/Friday flipped students, especially with respect to active learning, an atmosphere conducive to learning, being satisfied with the instruction in the course, and recommending the professor. These findings suggest that while there does not appear to be a significant difference on student learning outcomes, students may not enjoy the flipped format as much during a Tuesday/Thursday class time. Unfortunately, why students in the Tuesday/Thursday flipped classes have more negative views remains unknown. The negative views may be due to the nature of the flip itself and having to watch “more” videos between classes or due to the structure of the in-class learning activities. For example, in the Monday/Wednesday/Friday flipped

Table 6
Effect of Flipping on Student Evaluations, All Classes.

	Active Learning	Manage Time	Atmosphere	Good Assignments	Learn A Lot	Satisfied with Quality	Recommend Professor
Flip (=1)	0.814***	−0.296	0.385	0.014	0.211	0.408	0.800**
	(0.272)	(0.302)	(0.285)	(0.336)	(0.309)	(0.341)	(0.374)
Tues/Thurs Class (=1)	0.924***	0.521**	0.754***	0.820***	0.471	0.864***	0.958***
	(0.238)	(0.204)	(0.254)	(0.275)	(0.309)	(0.301)	(0.340)
Flip*Tues/Thurs Class (=1)	−1.195**	−0.265	−0.945*	−0.519	−0.849	−1.189**	−1.943***
	(0.472)	(0.517)	(0.520)	(0.524)	(0.549)	(0.579)	(0.606)
Online Evaluation (=1)	0.360	0.126	0.122	0.532	0.449	0.319	0.925*
	(0.457)	(0.523)	(0.515)	(0.468)	(0.473)	(0.527)	(0.541)
Constant	5.370***	5.822***	5.589***	5.380***	5.500***	5.250***	5.042***
	(0.190)	(0.157)	(0.194)	(0.210)	(0.190)	(0.214)	(0.238)
Observations	190	191	190	189	189	189	189
Adjusted R-squared	0.059	0.012	0.012	0.032	−0.004	0.015	0.039

Robust standard errors in parentheses. *** p < .01, ** p < .05, * p < .10.

Table 7
Student evaluations, flipped classes only, by class meetings.

	Tuesday/Thursday			Monday/Wednesday/Friday			Difference		
	Mean	Std. Err.	n	Mean	Std. Err	n	Mean	t-Stat	N
Active Learning	5.86	0.383	22	6.18	0.232	38	−0.32	−0.84	60
Manage Time	5.73	0.464	22	5.53	0.281	38	0.20	0.43	60
Atmosphere	5.73	0.421	22	5.97	0.255	38	−0.25	−0.59	60
Good Assignments	5.64	0.448	22	5.39	0.271	38	0.24	0.54	60
Learn A Lot	5.25	0.442	20	5.71	0.260	38	−0.46	−1.04	58
Satisfied w/Quality	5.25	0.482	20	5.66	0.283	38	−0.41	−0.85	58
Recommend Prof	4.90	0.507	20	5.84	0.298	38	−0.94	−1.86	58

classes, Fridays were reserved for completing the online problem sets, and as a result, students attended two classes in between the problem set days where they engaged in different active learning activities. Meanwhile, in the Tuesday/Thursday classes, roughly 55 min of the Thursday class was reserved for the online problem sets, so every other class period students completed the online work, leaving only one day per week of completely new and different activities. This may explain why students in the Tuesday/Thursday flipped classes ranked active learning lower than in the Monday/Wednesday/Friday flipped classes, if they do not view working together on online problem sets to be active learning but rather consider more physically active experiments, where they physically leave their seats to demonstrate something such as trading goods or producing goods, to be “active” learning. Furthermore, if students do not like the structure of a course, they will be less likely to like the quality of instruction and less likely to recommend the professor. Discerning the mechanism behind the differences in results is not possible given the current dataset.

Finally, those who filled out the online evaluation were significantly more likely to recommend the professor (at the 10% level) than those who filled out a paper evaluation, though there were no other significant differences. A common concern about online evaluations is that they are likely to suffer from a selection bias, as students elect whether or not to respond and must log in to do complete the online evaluation, whereas those who fill out the paper surveys typically are administered the survey during class, decreasing the marginal cost of filling out an evaluation compared to the online format.

Because the nature of the evaluation changed from a paper to online format, and because over the course of all the semesters course preparation materials changed, course evaluations may reflect these changes. During the Fall 2013–Spring 2014 academic year, all course preparation materials, including the online problem set platform, reading guides, and worksheets were identical between semesters. Additionally, during this academic year, all classes completed paper evaluations, diminishing the selection effect associated with filling out an online survey. Furthermore, the instructor taught flipped classes on Monday, Wednesday, Friday during the Fall semester and Tuesday/Thursday during the Spring semester. Therefore, a simple difference in means test among students in the flipped classes for each semester indicates if students view the flipping differently depending on the day of the week the course is taught. The results are presented in Table 7.

Table 7 indicates that for the majority of the questions, those in the Tuesday/Thursday classes rated the professor and class lower, though these differences are not significant, except for if the student would recommend the professor. In this case, those in the Tuesday/Thursday class are significantly less likely to report that they would recommend the professor, similar to the findings presented in Table 6. Unfortunately, one cannot discern if the difference in recommending the professor is because students did not like the nature of the flip itself, if it is because they were in smaller classes, or if it was a combination of the two: being in a small, flipped class. The instructor personally believes that the flip was not as satisfying to students in the smaller class setting, based on observational data on how smoothly activities progressed and how the students seemed to enjoy the activities compared to other flipped classes.

5. Discussion

This study documents positive and significant effects of flipping on students’ final exams compared to traditionally taught classes. It further finds that the effect of flipping does not vary by class scheduling, or the frequency and duration of the class, nor does it vary by class size. While there was no significant difference of the effect of flipping on student outcomes by class scheduling, those who met in a flipped setting less frequently generally had more negative views of the course as they reported less satisfaction with respect to active learning, an atmosphere conducive to learning, being satisfied with the instruction in the course, and recommending the professor. Therefore, while the effect of flipping may not differ with scheduling, the negative impact of flipping on student attitudes in classes that meet less frequently may be worthwhile for instructors to consider when scheduling their classes.

As this study was a retrospective study from a convenience sample, the data in this study are not perfect. First, the student learning outcome variables are midterm and final examinations, which are not identical from semester to semester. While the exam structure is the same, and while the instructor believes the difficulty is similar from semester to semester, there may be unintended

variation in exam difficulty. Furthermore, exams may also unintentionally weight topics differentially, which could also impact outcomes. However, these effects may be negligible as there is no reason to believe the exams varied in difficulty or content systematically from one semester to another. An advantage to having different exams is that the probability of and concern about cheating is eliminated, especially at a small college where students may have a greater ability to share resources from semester to semester.

Another limitation in this analysis is that course materials changed over time, as previously mentioned. While the lecture note base was the same, specific examples may have been updated to be more interesting, relevant, or timely, which perhaps may help students to remember the examples and/or the material. Some activities in the flipped classes were eliminated, altered, or added over time based on instructor perception of their successfulness or student feedback. Some students also had an advantage of having reading guides to help with reading the text or organizing their notes. The instructor specifically switched online problem set platforms due to student complaints, so if students preferred one platform to another, that may also have impacted student effort and learning.

While this study specifically aims to discern if the effect of flipping and student evaluations of flipping differ for classes that meet three times per week for 55 min versus twice per week for 80 min, one should note that students typically select their own class schedule. Therefore, there may be differences with respect to the types of students who may enroll in classes on the different days. For example, students may select their schedules either to commute to campus fewer days per week, to accommodate outside jobs or sports schedules, or to accommodate another unobservable characteristic. Therefore, results should be interpreted with caution due to selection issues.

There were only two classes that had fewer than thirty students in them, and the larger classes had a maximum of 39 students. Therefore, while this study attempts to document the effect of flipping by class size, the variability in class size is limited. Furthermore, there was one semester where the instructor did not take attendance, and this semester coincided with the small flipped classes, making it impossible to control for class size and attendance at the same time.

While this study also attempts to investigate how students in the different classes may perceive flipping, results involving student evaluations should be interpreted with caution. First, linking evaluations to student academic and demographic outcomes is impossible, limiting the ability to extrapolate and interpret much information from the evaluations. Furthermore, evaluations may have reflected the changing material and resources available to students, not just the class format. Additionally, as mentioned, the format of the evaluation changed from a paper format to an online format, resulting in greater selection bias. During the paper evaluation period, all students who were present in the class were provided with evaluations. The marginal cost of filling out an evaluation was therefore small. When the system switched to an online format, students could easily elect not to fill out the evaluation, as they simply did not log in to fill them out. The marginal cost of filling out evaluations was subsequently larger for the online format. Therefore, the online evaluations suffer from a greater selection bias than the paper ones.

The data in this study are from a small, private college setting with the same professor. Given the size of the college and the size of the classes, as well as the diversity of the student body, the generalizability of these findings to a larger, public university remains unclear.

Finally, while this study identifies and confirms a similar effect of flipping as previously documented using a larger sample size over multiple years, this study is still unable to identify the mechanism through which flipping is effective.

6. Conclusions

Flipping the class is a popular technique in higher education, and recent studies have documented flipping to improve student-learning outcomes in the principles of microeconomics classroom. Researchers have yet to uncover the mechanism through which flipping is successful, nor have they identified if the effectiveness of flipping varies by class spacing. A benefit of the flipped class is that it allows students to engage in active learning during class time either individually, in pairs, or in groups. If the classes are not long enough, activities may be interrupted and have to be resumed the following class period, perhaps weakening their effectiveness. Thus, longer class periods may allow for activities to be completed in one class period only. On the other hand, having to resume the activity another day may require students to revisit the material and refresh their memories on the activity, thus reinforcing the material.

Studies have also yet to identify if class size affects the flipping effectiveness. As students may often work in groups during class time, smaller classes may not allow for a large enough quantity of groups to make activities effective, large enough groups, or the ability to change groups, all of which may benefit the student or enhance the activities. On the other hand, being in a smaller class may make the course more intimate, allowing students to feel more comfortable and at ease with one another, encouraging learning, teamwork, communication, and cooperation.

This study utilizes a retrospective dataset of one instructor's principles of microeconomics classes over the course of five semesters to identify the effect of flipping on student learning outcomes, as well as how the effect may differ depending on class duration and frequency and class size. It further investigates how student evaluations may differ along these same characteristics.

Students in all classes had the same lecture notes base, with students in the traditional classes experiencing the notes live in class, and those in the flipped classes receiving the notes via online videos. All classes also completed online problem sets, though students in the flipped class were able to work in groups on the problems sets in class while those in the traditional class had to complete the

problem sets as homework outside of class. All classes had most of the same worksheets, though those in the flipped classes had additional worksheets corresponding to active learning activities conducted during the class period. Exams were not identical across classes, though they were similar in structure and content. At the end of the semester, students completed course evaluations in either paper or online format.

Using a larger data set over a longer time period, this study identified a similar effect of flipping on student outcomes as had previously been documented in the literature, reinforcing the current findings. In particular, it illustrated that students in the flipped class setting scored 0.216–0.324 standard deviations, or roughly one-third to one-half of a letter grade, higher on final exams compared to those in a traditional class. Class scheduling had no significant effect on exam scores. Among those in the flipped class, class size also did not impact learning outcomes.

This study next investigated how students perceived the course. Students in the flipped class setting that met three days per week rated the instructor significantly higher in encouraging active learning than those in the traditional Monday/Wednesday/Friday classes. They also rated the instructor higher in that they would recommend the instructor to another student, something tenure-track professors may value greatly. Students meeting on Tuesday/Thursdays, less frequently for a longer duration, also rated the instructor higher than those meeting on Monday/Wednesday/Friday with respect to the instructor encouraging active learning, managing class time effectively, and providing an atmosphere conducive to learning in the traditional classes. They further reported higher evaluations with respect to assignments helping them learn the material, being satisfied with the quality of instruction in the course, and recommending the professor to another student. Alternatively, students in the flipped classes on Tuesday/Thursday, however, were less satisfied and reported lower evaluations than those in the Monday/Wednesday/Friday flipped class.

Therefore, flipping has a positive and significant impact on student final exam scores compared to those in a traditional class, and the spacing meetings does not affect exam scores. Students in the Monday/Wednesday/Friday flipped class generally provided more positive feedback in course evaluations than those in the Monday/Wednesday/Friday traditional classes, indicating that students may enjoy the flipped format more, which may be part of the mechanism through which flipping is effective. However, students who experienced the flip less frequently for longer periods generally had more negative views of the course overall. Therefore, the results documented in this study suggest that instructors considering flipping their courses need not worry that scheduling will impact the effectiveness of flipping on student learning outcomes; however, students appear to respond more favorably to flipped courses which meet more frequently for shorter periods versus less frequently for longer periods. Therefore, when considering the scheduling of a flipped course, instructors should carefully weigh and consider the importance of student satisfaction for themselves, their students, department, and school.

Acknowledgements

I would like to thank Gary Maggs for encouraging me to pursue new pedagogies in my class and to conduct research on their effectiveness. My gratitude to Katherine Sauer for her mentorship in flipping and teaching. I am grateful to Ben Niu for insightful comments. I thank two anonymous referees for helpful comments and suggestions. I am thankful to Katie Sabourin and OIT for technical assistance and to St. John Fisher College for providing summer professional development support and opportunities. Liz LaChance provided excellent assistance with data retrieval.

Appendix A

Tables A1–A5 present the results of the analysis using cluster-robust standard errors, where standard errors are clustered at the class level. Results should be interpreted with caution due to the small number of clusters.

Table A1 corresponds to Table 2 in the main text. The main variable of interest, flipping, has increased in significance in Model I, with those in the flipped course scoring 0.324 standard deviations higher than those in the traditional class on the final exam in Model I (significant at the 1% level) but has become insignificant in Model II.

Table A2 corresponds to Table 3 in the main text and estimates Eq. (2) to determine if the effect of flipping varies by scheduling. Flipping is now significant at the 10% and 5% levels on the final exam in Models III and IV, and scheduling appears to have an effect. Students in traditional Tuesday/Thursday courses appear to perform significantly better than those in traditional Monday/Wednesday/Friday courses on the first midterm versus the second in Models III and IV, though the significance disappears in Model V. Finally, those in the flipped Tuesday/Thursday course appear to score better on Exam 1 in Model III (significant at the 10% level), though this significance disappears in Models IV and V. The interaction also becomes negative and significant (at the 10% level) for Exam 2 in Models IV and V, indicating that those in the flipped Tuesday/Thursday course perform worse on the second exam compared to those in the flipped Monday/Wednesday/Friday class. The effect disappears by the end of the semester and the final exam, suggesting any effect may be due to some sort of temporary adjustment period. This table provides suggestive evidence that the effect of flipping may vary with scheduling, though these results should be interpreted with caution due to endogeneity of controls and the small number of clusters.

Table A3 presents the results of the parsimonious estimation of Eq. (2) and corresponds to Table 4. The significance of the effect of flipping for the Monday/Wednesday/Friday class on the final exam has increased for all specifications, and the effect of scheduling on the test scores has also become significant. Specifically, those in a traditional Tuesday/Thursday course appear to do significantly better on the first exam across all specifications, but significantly worse on the final exam compared to those in the traditional Monday/Wednesday/Friday classes. There does not appear to be a significant effect on the second exam. This finding suggests that in a traditional class, students in the Tuesday/Thursday courses may fall behind midsemester as the demands of all classes, along with

Table A1
OLS Analysis of the effect of flipping on student outcomes, all classes.

	Model I			Model II		
	Exam 1	Exam 2	Final	Exam 1	Exam 2	Final
Flip (=1)	0.081	0.070	0.324***	0.057	−0.002	0.216
	(0.112)	(0.134)	(0.083)	(0.109)	(0.184)	(0.118)
Absences				−0.119**	−0.185***	−0.178**
				(0.036)	(0.038)	(0.054)
Senior (=1)	0.534**	0.699**	1.331***	0.481	0.508	0.693**
	(0.208)	(0.247)	(0.283)	(0.392)	(0.298)	(0.245)
Junior (=1)	0.226	0.052	0.062	0.102	−0.005	0.279
	(0.477)	(0.626)	(0.619)	(0.582)	(0.746)	(0.570)
Sophomore (=1)	0.330**	0.017	−0.099	0.305*	−0.068	−0.001
	(0.125)	(0.113)	(0.105)	(0.116)	(0.153)	(0.096)
AP or College Credit (=1)	−0.010	0.053	0.183	−0.058	−0.017	0.155
	(0.114)	(0.118)	(0.133)	(0.126)	(0.108)	(0.147)
AP Econ Credit (=1)	−0.041	0.227	0.254	−0.059	0.209	0.119
	(0.186)	(0.144)	(0.260)	(0.177)	(0.131)	(0.210)
SAT/ACT Convert Score (=1)	0.005***	0.004***	0.004***	0.005***	0.004***	0.004***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)
Commuter (=1)	0.207**	0.109	0.200	0.190	0.086	0.133
	(0.072)	(0.116)	(0.162)	(0.093)	(0.147)	(0.201)
First Generation Scholar (=1)	0.439	0.405	0.148	0.471	0.406	0.194
	(0.278)	(0.217)	(0.234)	(0.315)	(0.233)	(0.183)
Female (=1)	0.200	0.096	0.308*	0.207	0.063	0.220
	(0.178)	(0.079)	(0.132)	(0.182)	(0.039)	(0.134)
Family & Demographic Controls	Y	Y	Y	Y	Y	Y
Constant	−2.856	−1.172	−0.834	−4.919	−4.193	−2.150
	(2.009)	(1.586)	(2.232)	(2.920)	(2.294)	(2.076)
Observations	164	163	166	148	149	150
Adjusted R-squared	0.363	0.222	0.295	0.358	0.282	0.346

Cluster-robust standard errors in parentheses. *** p < .01, ** p < .05, * p < .10.

Table A2
OLS Analysis of the effect of flipping & scheduling on student outcomes, all classes.

	Model III			Model IV			Model V		
	Exam 1	Exam 2	Final	Exam 1	Exam 2	Final	Exam 1	Exam 2	Final
Flip (=1)	0.049	0.154	0.261*	0.014	0.180	0.270**	0.013	0.147	0.233
	(0.146)	(0.213)	(0.126)	(0.082)	(0.105)	(0.108)	(0.099)	(0.136)	(0.124)
Tues/Thurs (=1)	0.377**	0.062	−0.092	0.207*	−0.091	−0.145	0.237	−0.043	−0.091
	(0.146)	(0.213)	(0.126)	(0.100)	(0.123)	(0.136)	(0.131)	(0.188)	(0.186)
Flip*Tues/Thurs (=1)	0.302*	−0.020	0.324	0.127	−0.264*	0.175	0.090	−0.434*	−0.034
	(0.149)	(0.239)	(0.193)	(0.164)	(0.124)	(0.149)	(0.141)	(0.171)	(0.147)
Absences							−0.117**	−0.193***	−0.179**
							(0.038)	(0.038)	(0.054)
Academic controls	N	N	N	Y	Y	Y	Y	Y	Y
Family & Personal Demographic Controls	N	N	N	Y	Y	Y	Y	Y	Y
Observations	164	163	166	164	163	166	148	149	150
Adjusted R-squared	0.092	−0.010	0.039	0.365	0.221	0.288	0.362	0.288	0.338
F-stat	65.71***	1.02	10.21**	0.54	2.57	11.80***	0.43	5.06*	3.43

Cluster robust standard errors in parentheses. *** p < .01, ** p < .05, * p < .10. F-test is for joint significance of flipping variables. Academic controls include class year, AP credit, AP econ credit, SAT & ACT scores, and commuter and first generation status. Family & personal demographic controls include sex, age, ethnicity, household income, and parental education levels.

extracurricular activities, increase, perhaps because it is more difficult to keep up with the material when only meeting twice a week compared to three times per week. The scheduling of Tuesday/Thursday courses also means that students may go 4 days without thinking about the material, whereas in the Monday/Wednesday/Friday classes, students are forced to think about the material after only a two-day weekend. The effect of flipping does not appear to vary with scheduling for nearly all specifications, except for on Exam 2 in model P1, where those in the flipped Tuesday/Thursday courses score lower on the second exam compared to those in the flipped Monday/Wednesday/Friday classes.

Table A4 corresponds to Table 5 and presents the effect of scheduling and class size in the flipped classes only. Here, scheduling again has an effect on test scores, with students in the Tuesday/Thursday course performing significantly worse on Exam 2 in both specifications, but scoring significantly better on Exam 1 in Model FP2. There does not appear to be an effect on the final exam. Class

Table A3
OLS analysis of the effect of flipping & scheduling on student outcomes, parsimonious specifications, all classes.

	P1			P2			P3		
	Exam 1	Exam 2	Final	Exam 1	Exam 2	Final	Exam 1	Exam 2	Final
Flip (=1)	0.035	0.195	0.326***	0.030	0.169	0.302**	0.046	0.150	0.300***
	(0.068)	(0.107)	(0.082)	(0.093)	(0.136)	(0.100)	(0.092)	(0.145)	(0.058)
Tues/Thurs (=1)	0.187**	−0.085	−0.189*	0.190*	−0.097	−0.206*	0.210*	−0.080	−0.199**
	(0.064)	(0.104)	(0.078)	(0.092)	(0.139)	(0.102)	(0.101)	(0.148)	(0.081)
Flip*Tues/Thurs (=1)	0.067	−0.304*	0.063	0.076	−0.221	0.153	0.029	−0.221	0.145
	(0.169)	(0.146)	(0.172)	(0.149)	(0.141)	(0.141)	(0.127)	(0.169)	(0.109)
Class Year & SAT/ACT Scores	Y	Y	Y	Y	Y	Y	Y	Y	Y
Personal Demographic Controls	Y	Y	Y	N	N	N	N	N	N
Family Demographic Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	164	163	166	164	163	166	164	163	166
Adjusted R-squared	0.369	0.232	0.283	0.359	0.209	0.280	0.371	0.207	0.290
F-test	0.46	2.24	14.41***	0.69	3.35	19.93***	0.58	1.11	16.99***

Robust standard errors in parentheses. *** p < .01, ** p < .05, * p < .10. F-test is for joint significance of flipping variables. All models include controls for sex. Personal demographic controls include age, ethnicity, and first generation status. Family demographic controls include family income and parental education levels.

Table A4
OLS Analysis of scheduling & class size on student outcomes, flipped classes only.

	FP1			FP2		
	Exam 1	Exam 2	Final	Exam 1	Exam 2	Final
Tues/Thurs (=1)	0.205	−0.481*	−0.015	0.256**	−0.429**	0.072
	(0.112)	(0.186)	(0.142)	(0.059)	(0.104)	(0.103)
Small Class (=1)	−0.170	0.079	−0.094	−0.116	0.147	−0.002
	(0.124)	(0.204)	(0.290)	(0.133)	(0.181)	(0.228)
First Generation Scholar (=1)				0.431*	0.454	0.732*
				(0.142)	(0.242)	(0.245)
Class Year & SAT/ACT Scores	Y	Y	Y	Y	Y	Y
Family & Personal Demographic Controls	Y	Y	Y	Y	Y	Y
Observations	73	71	73	73	71	73
Adjusted R-squared	0.338	0.127	0.163	0.349	0.133	0.205

Cluster-robust standard errors in parentheses. ***p < .01, **p < .05, *p < .10. Family & personal demographic controls include sex, age, ethnicity, family income, and parental education levels. FP3 includes controls for first generation scholar status.

Table A5
Effect of Flipping on Student Evaluations, All Classes.

	Active Learning	Manage Time	Atmosphere	Good Assignments	Learn A Lot	Satisfied with Quality	Recommend Professor
Flip (=1)	0.814**	−0.296	0.385	0.014	0.211	0.408	0.800*
	(0.311)	(0.199)	(0.398)	(0.333)	(0.172)	(0.322)	(0.397)
Tues/Thurs Class (=1)	0.924**	0.521**	0.754	0.820**	0.471**	0.864**	0.958*
	(0.311)	(0.199)	(0.398)	(0.333)	(0.172)	(0.322)	(0.397)
Flip*Tues/Thurs Class (=1)	−1.195**	−0.265	−0.945*	−0.519	−0.849***	−1.189**	−1.943**
	(0.396)	(0.199)	(0.422)	(0.354)	(0.198)	(0.453)	(0.563)
Online Evaluation (=1)	0.360	0.126***	0.122	0.532***	0.449***	0.319	0.925*
	(0.245)	(0.009)	(0.143)	(0.120)	(0.098)	(0.319)	(0.399)
Constant	5.370***	5.822***	5.589***	5.380***	5.500***	5.250***	5.042***
	(0.311)	(0.199)	(0.398)	(0.333)	(0.172)	(0.322)	(0.397)
Observations	190	191	190	189	189	189	189
Adjusted R-squared	0.059	0.012	0.012	0.032	−0.004	0.015	0.039

Cluster-robust standard errors in parentheses. ***p < .01, ** p < .05, * p < .1.

size also does not have any effect on exam scores. Therefore, these results suggest that those in the Tuesday/Thursday courses may struggle to keep up with the material as the semester progresses and demands of other classes increase, as they are not forced to check in with the material as frequently as those in a Monday/Wednesday/Friday course; however, after an adjustment period, there does not appear to be a significant effect of spacing on flipping by the end of the semester on the final exam, suggesting any effect may be temporary.

Finally, Table A5 corresponds to Table 6 and presents the effect of flipping and scheduling on student course evaluations. Notable

differences between Tables A5 and 6 include that those in the traditional Tuesday/Thursday class are now significantly more likely to report they learned a lot but there is no difference in their view of the learning atmosphere compared to those in a traditional Monday/Wednesday/Friday class. Those in the flipped Tuesday/Thursday class report significantly lower rankings in the category of believing they learned a lot. Interestingly, those students who filled out the online evaluations appear significantly more satisfied with how the instructor manages time, thinking that the assignments helped them to learn the material, and that they learned a lot in the course. These results may indicate a positive selection bias in the online evaluations, though it is difficult to tell if it is selection bias or if it is because the online evaluations correspond to a period in time after the instructor had taught the course for multiple semesters and may have improved the course overall to address previous students' concerns.

Thus, the cluster-robust specifications generally support the main findings in that students in a flipped Monday/Wednesday/Friday class scored significantly higher on the final exam compared to those in the traditional Monday/Wednesday/Friday class; however, the effect of flipping may vary with scheduling as it may impact students' midterm exam scores. The differential impact of the flip with respect to scheduling appears to be short-term, as there is no significant effect by the end of the semester, suggesting the class format may take a little longer for students to adjust to in the class that meets less frequently. These results should be interpreted with caution as there are a small number of clusters and the significance of the impact appears to vary with model specification. Scheduling also may impact students' evaluation and views of the course overall, something those on the tenure track or contract instructors may with to consider when scheduling and deciding whether or not to flip their classes. Future research could address whether these findings are due to a small number of clusters or if they are in fact due to the impact of scheduling by investigating a larger dataset.

References

Arias, J., Walker, D.M., 2004. Additional evidence on the relationship between class size and student performance. J. Econ. Educ. 35 (4), 311–329.
Balaban, R.A., Gilleskie, D.B., Tran, U., 2016. A quantitative evaluation of the flipped classroom in a large lecture principles of economics course. J. Econ. Educ. 47 (4), 269–287.
Bergman, J., Sams, A., 2012. Flip Your Classroom: Reach Every Student in Every Class Every Day. International Society for Technology in Education, Eugene, Or.
Calimeris, L., Sauer, K., 2015. Flipping out about the flip: all hype or is there hope. Int. J. Econ. Educ. 13–28.
Carrington, L.G., Houston, S., 2010. The impact of course scheduling on student success in intermeidate accounting. Am. J. Bus. Educ. 3 (4), 51–60.
Caviglia-Harris, J., 2016. Flipping the undergraduate economics classroom: using online videos to enhance teaching and learning. South. Econ. J. 83 (1), 321–331.
Daniel, E.L., 2000. A review of time-shortened courses across disciplines. Coll. Stud. J. 34 (2), 298–306.
Dempster, F.N., 1988. The spacing effect: a case of history in the failure to apply the results of pyschological research. Am. Pyschol. 43 (8), 627–634.
Diette, T.M., Raghav, M., 2015. Class size matters: heterogeneous effects of larger classes on college student learning. East. Econ. J. 41, 273–283.
Gallo, M.A., Odu, M., 2009. Examining the relationship between class scheduling and student achievement in college algebra. Commun. Coll. Rev. 36 (4), 299–325.
Green, A., 2014. The case for the traditional classroom. Int. Rev. Econ. Educ. 16 (Part B), 87–99.
Henebry, K., 1997. The impact of class schedule on student performance in a financial management course. J. Educ. Bus. 73 (2), 114–120.
Hughes, H., 2012. Introduction to flipping the college classroom. In: Amiel, T., Wilson, B. (Eds.), Proceedings of World Conference on Educational Multimedia, Hypermedia, and Telecommunications. Chesapeake, VA: AACE. pp. 2434–24380.
Lage, M., Platt, G.J., Treglia, M., 2000. Inverting the classroom: a gateway to creating an inclusigve learning environment. J. Econ. Educ. 31 (1), 30–43.
Larson, S., Yamamoto, J., 2013. Flipping the college spreadsheet skills classroom: initial empirical results. J. Emerg. Trends Comput. Inform. Sci. 4 (10), 751–758.
Miyamoto, Y.R., Coleman, C.A., 2015. Beyond time-on-task: the relationship between spaced study and certification in MOOCs. J. Learn. Anal. 2 (2), 47–69.
Monks, J., Schmidt, R.M., 2011. The impact of class size on outcomes in higher education. BE J. Econ. Anal. Policy 11 (1), 1–17.
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 Sci. Educ. 9, 473–481.
O'Flaherty, J., Phillips, C., 2015. The use of flipped classrooms in higher education: a scoping review. Internet High. Educ. 25, 85–95.
Olitsky, N.H., Cosgrove, S.B., 2016. The better blend? Flipping he principles of microeconomcis classroom. Int. Rev. Econ. Educ. 21, 1–11.
Pierce, R., Fox, J., 2012. Vodcasts and active-learning exercises in a flipped classroom model of a renal pharmacotherapy module. Am. J. Pharm. Educ. 76 (10), 196.
Roach, T., 2014. Student perceptions toward flipped learning: new methods to increase interaction and active learning in economics. Int. Rev. Econ. Educ. 17, 74–84.
Scott, P.A., Conrad, C.F., 1991. A Critique of Intensive Courses and an Agenda for Research. Division of Summer Session and Inter-college Programs, University of Wisconsin, Madison ERIC Document Reproduction Service No. ED337087.
Trueman, M., Hartley, J., 1996. A comparison between the time-management skills and academic performance of mature and traditional-entry university students. High. Educ. 32 (2), 199–215.