2019

Learning Effects of the Flipped Classroom in a Principles of Microeconomics Course Running Header: Flipped Principles of Micro

Erik Craft
University of Richmond, ecraft@richmond.edu

Maia K. Linask
University of Richmond, mlinask@richmond.edu

This is a pre-publication author manuscript of the final, published article.

Recommended Citation

Craft, E., & Linask, M. (2019). Learning effects of the flipped classroom in a principles of microeconomics course. *The Journal of Economic Education*, 1-18. https://doi.org/10.1080/00220485.2019.1687372

Learning Effects of the Flipped Classroom in a Principles of Microeconomics Course
Running Header: Flipped Principles of Micro

Erik Craft
Maia Linask

August 2019

Erik Craft and Maia Linask are associate professors of economics in the Robins School of Business at the University of Richmond, VA 23173. Their telephone numbers are (804) 287-6573 and (804) 287-6520, and their email addresses are ecraft@richmond.edu and mlinask@richmond.edu, respectively. Their mail address is Department of Economics, Robins School of Business, 102 UR Drive, University of Richmond, VA 23173. Fax number is (804) 289-8878. Craft is the corresponding author.

We thank the participants of a Conference on Teaching Research in Economic Education (CTREE) session in Atlanta in May/June 2016, a 2017 University of Richmond Economics seminar, James Hornsten and others at the 2018 Western Economic Association Annual Conference, and three anonymous referees for their helpful comments. We also acknowledge the assistance of the University of Richmond's Admissions Office, Registrar's Office, Financial Aid Office, and Office of Institutional Effectiveness. The University of Richmond Institutional Review Board approved the authors' research protocol for requesting student's personal data and its use.

Abstract: This paper estimates learning effects of the flipped classroom format using data from sixteen sections of Principles of Microeconomics over a four-year period. The experimental design is unique in that two treatment and two control sections were taught during the fall semester in four consecutive years. Further, the instructor switched the time of day when the treatment and control sections were taught each year. Controlling for gender, ACT score, a normed high school GPA, Pell Grant award, time of day, and initial knowledge of economics, the study finds no evidence of increased learning using end-of-semester measures for students in the flipped classroom in comparision to sections with a moderate amount of active learning.
Key Words: flipped classroom, inverted classroom, active learning, principles of microeconomics, and pedagogy.
JEL: A2

A recent pedagogical innovation is known as the flipped classroom. While definitions vary, the flipped classroom is generally understood as moving traditional lecture instruction outside of the classroom and instead using class time for active-learning activities, which include problem-solving, discussion, and experiments (see, for example, Lage, Platt, and Treglia, 2000, for an early application in economics). This paper analyzes various learning outcomes for students in flipped Principles of Microeconomics sections relative to a more traditional course format with a moderate amount of active learning. Particular care is taken to control for student aptitudes, initial knowledge of economics, and the time of day the class meets. To our knowledge, this paper is the first research on the flipped classroom in economics courses that both compares flipped and traditional course sections taught by the same instructor during the same semester and uses identical assignments and a standardized test of economic knowledge to measure both proficiency and growth. Our approach precludes misidentification of treatment effects due to year, cohort, time of day, or incentive differences.

Our paper seeks not only to identify whether learning is improved by a flipped classroom approach, but also to investigate whether the flipped classroom approach has different effects by gender, initial knowledge of economics, and overall academic preparation.

After four years and sixteen sections of Principles of Microeconomics (23-28 students per section), half of which were taught using a flipped classroom approach, we found no statistically significant end-of-semester treatment effects of the flipped classroom. While this result is at odds with much of the existing literature (see, for example, Balaban, Gilleskie, and Tran 2016 and Caviglia-Harris 2016), there is some overlap between the upper bound of our 95-percent confidence interval and the lower end of estimated impacts in the literature. Our results are consistent with existing evidence suggesting that the flipped classroom has a small impact on the final exam grade, which is the most directly comparable measure across studies: Calimeris (2018) finds that the flipped classroom increases the final exam grade by .324 standard deviations and Wozny, Balser, and Ives (2018) find that the flipped classroom increases final exam grades by .16 standard deviations for only the high-achieving students. While our point estimate of -0.761 (-0.04 standard deviations) is not statistically or economically significant, the upper bound of the 95-percent confidence interval (.17 standard deviations) is broadly consistent with these other findings. On the other hand, the lower bound of our confidence interval (-0.24 standard deviations) cannot rule out a negative effect of the flipped classroom.

Our study avoids a number of potential confounding factors. First, we compare results across class sections within the same semester. This eliminates the possibility that improved outcomes are due to improved teaching over time rather than from the flipped class format.[i] Furthermore, the students in both the flipped and the traditional classes had exactly the same assignments in our study and so were responding to the same incentives. Our paper thus contributes to the literature by running a rigorous, controlled teaching experiment over four semesters in which the only difference between the flipped and the traditional class with a moderate level of active learning pedagogy is the medium and timing of delivering new material via lecture and the activities to which class time is dedicated, i.e. the only difference is that in-class and out-of-class activities are flipped. We find that the point estimates of any impact of the flipped classroom are not statistically significant and that, at most, any positive impact of the flipped format is small (the upper bound on the 95-percent confidence interval is .13 to .36 standard deviations, depending on the exact outcome measure).

LITERATURE REVIEW

Economics pedagogy has been influenced by a long trend of exploring alternative methods of engaging students in their subject matter. Instructors have used classroom experiments for decades, and many other active learning techniques followed. While space constraints prevent a complete review of these developments, significant markers of the desire by economists to add variety and pedagogical innovations to instruction can be found in books by Becker and Watts (1998) and Becker, Watts, and Becker (2006) as well as the Handbook by Hoyt and McGoldrick (2012).

Recently, researchers in economics education have turned their attention to the flipped classroom and its potential to assist students in becoming more proficient in the development and application of their economics knowledge.[ii] In higher education, Crouch and Mazur (2001) are often cited as early practitioners of a form of the flipped classroom in physics instruction. Early papers in the economics education literature have discussed how to flip a classroom (Vasquez and Chiang 2015) and how students perceive flipped learning (Roach 2014), and studies on the impact of a flipped classroom have been carried out in virtually every discipline (for example, see Giannakos, Krogstie, and Chrisochoides, 2014 for a review of flipped classroom literature in computer science and Betihavas et al., 2016 for a review of the literature in nursing). While some papers find a positive effect of the flipped classroom on learning outcomes (e.g., Van Sickle 2016 in psychology) and others find no effect (e.g., Psihountas 2018 in finance), we have not been able to identify another paper that controls for as many confounding factors as possible by collecting data across multiple years, across multiple sections (in different formats) in the same year, and from the beginning and end of the semester for the same student. Our paper aims to fill that gap.

More recent papers in economics have undertaken a rigorous estimation of learning outcomes from a flipped classroom approach. Olitsky and Cosgrove (2016) provide some evidence of improvements in learning in a flipped-blended course.[iii] Both Balaban, Gilleskie, and Tran (2016) and Calimeris (2018) find that student performance increased on the final examination in their flipped classroom treatment group. They also find, respectively, no different marginal effects for students with different characteristics or for classes of different size or duration. Lombardini, Lakkala, and Muukkonen (2018) also find that the flipped classroom improved performance on the final exam and, in addition, reduced the likelihood of failing or withdrawing from a class. Calimeris and

Sauer (2015) similarly find that students in a flipped classroom do better on the final examination as well as the second midterm examination, while Wozny, Balser, and Ives (2018) find that the flipped classroom improves performance only for students with above-median GPAs on the final exam but for all students on mid-term exams.

A number of caveats, however, apply to the results in these papers. In particular, in all cases only students who participated in the flipped classroom format had to complete short-term assessments, such as daily quizzes, on the material that was introduced via video outside of class.[iv] Because students in the traditional classroom format were not required to complete these assessments, it remains unclear whether the improved exam performance was due to the class structure and flipped format or simply to the incentives that short-term assessments created for students to complete their assigned pre-learning activities.[v] In our experiment, we abstract from the question of how much preparation students do outside the classroom (under the assumption that daily quizzes incentivize greater preparation) in order to focus on the question of what students do in the classroom as compared to outside the classroom. Furthermore, the problems solved in the flipped class format as well as the solutions were made available to students in the traditional class.

There is also the potential for assessments used as measures of learning to be inadvertently designed or graded to favor the flipped class format. While other papers have taken steps to avoid such bias, our use of the Test of Understanding in College Economics (TUCE) takes an additional step in this direction by taking at least one assessment instrument out of the hands of the instructor entirely (see also Lombardini, Lakkala, and Muukkonen 2018, who also use the TUCE). The TUCE, published by

National Council on Economic Education and created by a committee of economics instructors, is a standardized test and thus not subject to the bias of a particular instructor. Administering the TUCE at both the beginning and end of the semester also allows us to control for prior economic knowledge and to use both proficiency and growth outcome measures.

A third concern is that if the traditional class and the flipped class are taught in different semesters or years, then the effects of the flipped classroom may be confounded with improved teaching due to the accumulation of experience (as noted in Balaban, Gilleskie, and Tran 2016). While Caviglia-Harris (2016) avoids the problem of different incentives noted above, it is subject to this potential challenge.

Our paper addresses all three of these concerns while also including controls for student characteristics. Students in both class formats completed exactly the same assignments before class and therefore had exactly the same incentives to prepare for class. We take additional steps to avoid bias by using the TUCE and by giving all students access to the same materials (other than the online lectures). Finally, our experiment ran over four years, with both flipped and traditional classes taught at different times in each semester so that we are able to clearly distinguish any flipped classroom effects from time of day, cohort, or time trend effects.[vi]

EXPERIMENTAL DESIGN AND DATA DESCRIPTION

One of the authors has taught Principles of Microeconomics for over twenty years at the University of Richmond. During each of the fall semesters of 2014, 2015, 2016, and 2017, he taught four sections of Principles of Microeconomics. The four sections met for fifty minutes on Mondays, Wednesdays, and Fridays at 9am, noon, 1:30pm, and 3pm.

Each fall, two of the sections were taught in a traditional lecture style with about four classroom experiments, a couple of short small group exercises, one day of discussion, and a day and a half completing together a worksheet on tax incidence and deadweight loss . The other two sections were taught in an entirely flipped format. With the exception of the first two class meetings, class time in the flipped sections was spent solving problems, engaging in classroom experiments (about ten experiments), and discussion.

All four sections received identical text and supplemental reading assignments throughout the course. In addition, the flipped sections were assigned online versions of the instructor's lectures, recorded on an iPad using the ExplainEverything app (voice over graphs, images, and drawings). Rather than 45 minutes, the online lectures (broken into multiple pieces) averaged 35-40 minutes. The lectures are more efficiently presented, since one can stay close to the script and graph while on pause during the recording process and need not stop to answer student questions. Nearly all of the lectures were updated/rerecorded during the summer of 2016. Access to these lectures was available only by logging in to an online module with the student's user name and password; the online lectures were thus not readily available to students in the traditional sections.[vii]

In order to identify different learning outcomes for the treatment sections, all the sections in a particular year were assigned the same chapter quizzes and problem sets and took the same mid-term and final examinations. The final examination each year included the TUCE, additional multiple-choice questions, and short answer problem-solving questions (some numerical, some conceptual, and some factual). All four sections began each class day with a quiz of four multiple-choice questions. During most class days, the questions were randomly selected from a test bank and were identical for all four sections, but on some days, up to two of the questions for the treatment sections came from the online lectures. The purpose of the quizzes with questions from the online lectures was to encourage students to watch the lectures before class. During the first two and one-half years of the study, results for the daily quizzes were provided in Blackboard at the end of each day. Beginning half way through year three, the results were shared with each section immediately after taking the quiz.

To control for time of day effects, during 2014, the flipped classroom sections met at noon and 1:30pm.[viii] This was repeated during the fall of 2016. In 2015 and 2017, the flipped classroom sections met at 9am and 3pm, while the traditional sections met at noon and 1:30pm. We believe this is the first research design to compare the outcomes of flipped and traditional course sections in the same semester across multiple years. This is critical because otherwise the order of the experiment may matter. In particular, in many flipped classroom experimental designs the default is to offer the traditional class in the first year or semester and the flipped class subsequently (see, for example, Caviglia-Harris 2016). In this case, it is difficult to determine whether improved outcomes are due to the changed instructional approach or simply a more experienced and better prepared instructor. Further, our research design is unique in that it can control for both year and time of day effects, which allows us to rule out more favorable class times as a reason for increased learning.

Students registered for a section in the preceding spring or summer without knowing which teaching format would be used. In fact, students likely had no idea that a pedagogical experiment existed until the initial class meeting when the research project was explained and they were asked to consent to share past and future educational data.[ix]

In particular, following Institutional Review Board rules, the instructor asked students for their consent in allowing the researchers access to their SAT/ACT scores, their University of Richmond Admissions normed high school GPA (beginning with the 2015 cohort), any Pell grant awards (beginning with the 2015 cohort), and use of their scores in their course. At the end of class on the first day, the instructor administered the microeconomics portion of the Test for Understanding College Economics (TUCE) to all students. Only on the second day of class did the instructor assign students who were in the treatment sections their first online lectures; since students were not aware whether they were in a treatment or control section when giving consent, the teaching style did not influence consent decisions by design.[x] Given the few openings remaining in other Principles of Microeconomics sections and the very short administrative deadline for adding courses (1 week from the start of the semester), it would have been very difficult for a student to change sections at that point.

Each section included between 23 and 26 students regardless of the year. In 2014, perhaps six to ten students in total dropped the course during the first week; about half of these students did not know whether or not they were in a treatment section. In 2015, a more careful description of the experiment on the first day of class led to fewer than five persons withdrawing from the four sections, most of these withdrawals taking place before they learned whether they were in a treatment section. This pattern continued in years 2016 and 2017. Students in the treatment sections did not learn they were in treatment sections until the second day of class, when the instructor assigned the first online video. Students in the control sections would only know they were in control sections via communication with students in the treatment sections or after weeks when the topic somehow came up in class.

In the years 2015, 2016, and 2017, 147 students were registered in flipped sections, of which 117 granted consent (79.6%) to participate in the study. In the traditional sections, 148 were registered and 122 gave consent (82.4%). The primary reason students did not give consent to participate in the study was that they were still seventeen years old and could not legally grant consent. A few students were not present in class on the first day when the TUCE pre-test was administered and consent was requested according to the IRB-approved procedure. A few foreign or foreign exchange students did not consent to share their data. Students who did not or could not grant consent to share data for the study were not removed from their section. Selection effects are therefore not expected to be important.

In 2016 and 2017, answers to problem set questions and additional questions solved in the treatment sections were posted on Blackboard for students in all sections. Since a conventional economics course can post additional problems and solutions, it seemed appropriate to make such questions and solutions available to students in control sections. Questions for chapter MyEconLab quizzes were randomly chosen with each new text edition at the beginning of the semester and reused the subsequent year. Online lectures were updated as regular lectures changed year to year. Class attendance was very high across all sections.

Table One presents summary statistics of the data collected in years 2015, 2016, and 2017. We restrict ourselves to the last three years' data in our baseline analysis for multiple reasons. First, the data for 2014 do not include two of the control variables, adjusted high school GPA and Pell Grant awards. We also hypothesize that the first year of a new teaching technique is not likely to have reached its full potential effectiveness. Finally, we initially exclude the 2014 results because a higher number of students dropped the course, and this ensures that our results are robust to any possible selection bias. Including the initial year of teaching with the flipped classroom technique does not affect our results, although our controls are more limited. Data from students who were too young to grant consent or who chose not to grant consent are, naturally, excluded from our analysis. A nontrivial number of students (nearly 10 percent) claimed to be below the age of eighteen at the beginning of each semester and could not therefore give consent to be part of the study.

[Insert Table 1 about here]

We use five different measures of learning to estimate the impact of a flipped classroom format on learning: final course grade, final exam grade, the problem-solving questions grade, the post-course TUCE score, and the percentage of the gap closed between the initial TUCE score and the maximum possible score of 30. The course grade measures the overall performance in the course. In order to standardize across years, it is measured as a percentage of the total available points. The final exam grade identifies performance on the comprehensive exam taken during exam week at the end of the course. The maximum number of points is 200.[xi] Because classroom activities in the flipped class often focused on solving problems, we also isolate the percentage of correct answers on the problem-solving portions of the final exam. In order to assess learning, the TUCE was administered again as part of the final exam. The post-course TUCE score for each student is the number of questions answered correctly (out of 30). The gap variable measures what share of the gap between the TUCE pre-test score and a perfect score of thirty was closed by the end of the semester: $gap = \frac{post\ course\ score - pre\ course\ score}{30 - pre\ course\ score}$. Thus, our investigation seeks not only to evaluate the effect of the treatment on students' level of economic knowledge, but also to identify differences in amount learned (i.e., progress) across the control and treatment groups.

There was significant variation within all five outcome variables (course grade, final exam grade, problem-solving questions, post-course TUCE, and gap). The average final exam grade and course grade averages were similar across the three years.[xii] It is worth noting that while most students' TUCE scores improved over the semester, there was one student in 2016 who scored the same on the pre-course TUCE test and the post-course TUCE test. On average, students closed 56% of the gap between their pre-course TUCE score and a perfect score of 30; this ranged from closing 54% of the gap in 2016 to closing 59% of the gap in 2017. It is also interesting that while the correlation between the final exam grade and the course grade was high, as expected (coefficient of .90), the correlation between the post-course TUCE variable and the course grade was much lower (coefficient of .68). Furthermore, the correlation between our measure of growth (the percentage of the TUCE gap closed by the end of the semester) and other outcome variables was quite low: .59 coefficient for the final exam and .49 coefficient for the course grade.

The baseline analysis includes thirteen independent variables: Treatment, Gender, ACT score, TUCE pre-test score, Pell Grant award, adjusted high school GPA, indicators for taking the final during the first and last scheduled exam times, and year and time of day indicators.[xiii] The estimating equation is thus

$$Outcome_{ijt} = \alpha + \beta_1 Treatment_{ijt} + \beta_2 Female_{ijt} + \beta_3 ACT_{ijt} + \beta_4 TUCE\ pre-test_{ijt} + \beta_5 Pell\ Grant_{ijt} + \beta_6 High\ School\ GPA_{ijt} + \beta_7 First\ Exam_{ijt} + \beta_8 Last\ Exam_{ijt} + \gamma_1 9am_{ijt} + \gamma_2 1{:}30pm_{ijt} + \gamma_3 3pm_{ijt} + \rho_1 2015_{ijt} + \rho_2 2016_{ijt} + \varepsilon_{ijt}$$

where $Outcome_{ijt}$ is one of our outcome varaibles for student *i* in section *j* in year *t.*

The variable of primary interest is Treatment, which is a binary variable equal to 1 if a student was in a flipped classroom section and 0 otherwise. Female is also a binary variable, equal to 1 for women and 0 otherwise. For students who did not take the ACT, we converted the combined math and critical reasoning SAT scores into a composite ACT score using a concordance provided by the College Board. The pre-course TUCE reports the number of correct answers on the initial administration of the TUCE at the beginning of the semester. ACT and TUCE pre-course scores are included to control for both students' ability or academic preparation as well as their initial level of knowledge. We also include the students' University of Richmond normed high school GPA and the Pell Grant amount.[xiv] The GPA is reported on the standard 4-point scale and presumably captures some combination of innate talent and work habits. The Pell Grant amount is given in thousands of dollars. (The majority of students did not receive a Pell Grant; the Pell Grant amount for those students is set to 0.) Pell Grant recipients come from lower income backgrounds and perhaps also lower quality schools or schools that do not offer AP courses, so this variable controls at least partially for socio-economic background. Because the meeting time of a section might influence learning outcomes (maybe motivated students select early section times or perhaps early section times make learning difficult), we also control for section meeting times. The day of the final exam might affect performance on the exam, so we have also investigated the effect of an early or late exam day.[xv] Finally, Table One reports the statistics on the (self-reported) hours spent reading the text, working on MyEconLab quizzes, working on optional MyEconLab exercises, general study, and viewing online lectures.

A few points about the control variables are worthy of note. Just under one-half of each sample was part of the treatment group. Interestingly, only 34% of the whole sample were females. While the mean pre-course TUCE was significantly lower than the mean post-course TUCE as expected, the standard deviation was slightly greater in the pre-test. The average of the adjusted high school GPA was almost identical across the three years; the ACT score of the 2017 students was about .7 points higher than the previous two years. Indeed, the only other difference of note between 2015, 2016, and 2017 was when students took the final exam. In 2015, only 22% took the exam on the first available final exam date, and 39% took it on the last available final exam date. In 2016, a higher share of students took the exam on both the first available date (26%) and on the last available date (50%). In 2017, only 14.3% of the students took the exam on the first possible date, while 21.4% waited until the last day.[xvi]

We use two different estimators depending on the outcome variable under consideration. The course grade, the share of problem-solving questions correct, and the TUCE gap measure are all fractions between 0 and 1 (inclusive). We therefore use a fractional response model estimator with a loglog link function, which accounts for the absence of values near 0 for the course grade and the problem-solving questions share.[xvii] (Note that our tables report the marginal effect, not the coefficient estimate itself, in order to facilitate comparison between different outcome variables.) For the final exam grade and the TUCE post-test grade, we estimate a linear relationship. In both cases, because the flipped classroom treatement was assigned at the section-year level, we cluster standard errors at the section-year level (see Abadie et al., 2017 for an explanation).[xviii]

RESULTS

Table Two shows the baseline regression results with clustered standard errors at the section level for our five primary measures of learning. All twelve sections from 2015, 2016, and 2017 are included, totaling 237 students. Recall that not every student in the four sections is included in the data, mainly due to inability to give consent. A small number of students chose not to give consent.[xix] The results find no statistically significant effect of the flipped classroom approach for any of the five end-of-semester learning outcomes. Indeed, our estimates are quite close to zero in all five cases.[xx] The upper bound of the 95-percent confidence interval does allow the possibility of a small positive effect. Our estimates of the impact on the final exam grade are most directly comparable to other papers, and for this outcome the upper bound on the confidence interval is 1.6 percent, or less than two-tenths of a standard deviation. This is broadly consistent with the low-end estimates in the literature (see Wozny, Balser, and Ives 2018 and Calimeris 2018). On the other hand, the lower bound on the 95-percent confidence interval is -2.4 (almost one-fourth of a standard deviation), indicating that we cannot rule out a negative impact. For the other end-of-semester outcome variables, the confidence interval upper bound is .7 percent on the overall course grade (less than one-tenth of a standard deviation), 1.7 percent on the problem solving portion of the exam (.13 standard deviations), nearly 1 point on the TUCE test (.28 standard deviations), and 6.5 percent of the TUCE gap (.36 standard deviations). Thus, our point estimates are quite close to null and admit the possibility of, at most, a small positive effect. Naturally, the lower limit of the confidence interval also admits the possibility of a negative effect.

[Insert Table 2 about here]

The estimated effect of control variables is mostly of the expected sign. Higher ACT scores are statistically and economically significant in explaining learning outcomes. Each additional point on one's ACT score is associated with nearly an extra percent in the course grade, over two points on the final exam, and more than one percent higher score on the problem-solving portion of the exam. This is quite similar to the results found in Balaban, Gilleskie, and Tran (2016), where an additional point on the ACT raises the course grade by one-half of one percentage point; using the SAT instead of the ACT, Calimeris and Sauer (2015) also find that a higher standardized test score increases student performance on exams. Not surprisingly, a higher TUCE score at the beginning of the course statistically predicts a higher learning outcome at the end of the course.[xxi] For each $1000 awarded, Pell Grant recipients perform half a percentage higher in the course and 1.4 points (out of 200) higher on the final examination. They also scored .27 points higher on the end-of-semester TUCE test and closed 1.5 percent more of the gap between their pre-course TUCE score and the perfect score of 30. The UR adjusted high school GPA appears to positively affect the final course grade in a plausible manner. Going from an adjusted high school GPA of 3 to a GPA of 4 is expected to increase a student's final grade by nearly 5 percentage points. There is some evidence that students learn better in all sections other than the noon section, although this result is not consistent across all outcome variables. Students in the 3pm section are estimated to score five points more on the final examination relative to the noon section; this result is both statistically and economically significant.

Table Three reports the results of estimating equation (1) using graded work during the course of the semester as our measures of learning.[xxii] We find that the flipped classroom had a statistically significant impact in two cases: the first midterm and the online assignments. However, the effect was positive only in the case of the first midterm, while students in the flipped classroom actually performed worse on the online chapter quizzes in MyEconLab. The positive effect for the midterm is similar to the estimate in Wozny, Balser and Ives (2018), who find that the flipped class format has a positive impact in the medium term but not on long-term or end-of-semester outcomes. The flipped classroom treatment had no statistically significant impact on other assignments during the course of the semester (daily quizzes, problem sets, and second midterm) while the control variables mostly had the expected impact (ACT score, TUCE pre-test score, and high school GPA all had a positive relationship with grades during the semester). It is interesting to note that female students on average earned more points on the problem sets.

[Insert Table 3 about here]

Table Four reports the results of analyzing the self-reported number of hours spent reading the textbook, taking MyEconLab online quizzes, working on optional MyEconLab questions, studying in general, and working on microeconomics in total. The flipped classroom treatment has a statistically significant and positive effect on the total hours spent on microeconomics. Because the total hours spent on the course includes watching online lectures for students in the flipped class but not those in the traditional class, it is not surprising that students in the flipped class spent, on average, 3.1 hours more per week learning microeconomics. Indeed, this increase can be attributed almost entirely to the 2.7 average hours per week watching lectures that these students report.[xxiii] The academic preparation and aptitude variables suggest that students who are better prepared do not need to spend quite as much time studying. Finally, students for whom English is not a first language tend to spend more time on almost every type of studying.

[Insert Table 4 about here]

Three caveats for this self-reported data should be noted. First, students in the flipped classroom sections reported spending about two more hours per week on average when asked about time spent on various activities in a supplemental survey. But when reporting only total hours spent on the course in their official course evaluation (for which the authors have no individual identifiers), the average difference between the flipped sections and treatment sections was only one hour. This suggests that including additional categories of how one spent time led students in the treatment sections to bias upward their total time spent on the course. Hours spent on task might also be selecting for other student unobservables correlated with poor performance, in other words, the causation might be running from poor performance to more time spent studying.[xxiv] Finally, because students completed the survey at the end of the semester, their estimates may be biased upward by cramming at the end of the semester.

If the flipped classroom approach helps some students learn better and harms other students, real learning effects might exist but be unidentified in total population regressions. We repeated all of the above regressions with interaction variables between the treatment variable and various characteristics that might impact the effectiveness of a flipped classroom environment. For example, perhaps less able students learn better with a more active learning approach than do gifted students. Or perhaps students with less initial knowledge of economics perform better with an active learning approach, or maybe women and men respond differently to a flipped classroom approach. Or a flipped classroom environment might even work better during one part of the day than another.

Table Five gives the results, although we report only the coefficients for the treatment effect, the subgroup variable, and the interaction in order to conserve space. Most of these interaction terms were not statistically significant. Our analysis did identify a few statistically significant interaction effects. In particular, students in a 9am section scored lower on the final exam and the Post-TUCE exam as well as in the course as a whole when in a treatment section. Conversely, students in the 3pm flipped sections scored two points higher on those three outcomes. When interacting a student's ACT with the treatment variable, there is suggestive evidence that students with higher ACT scores tended to close more of the gap in the TUCE score between the beginning and end of the semester. The sign of the coefficient on the interaction terms for the other outcomes is also negative, although there is no statistically significant effect. This is consistent with the idea that students who are more prepared for college, *ceteris paribus,* are associated with greater effort or better study habits, which manifest themselves in a flipped classroom by watching the lecture videos more carefully or frequently.

[Insert Table 5 about here]

We find no evidence that women and men are affected differently by the flipped classroom. There is also no evidence that students with lower high school GPAs performed worse or better with a flipped classroom approach. The pre-course score on the TUCE, which controls for the level of knowledge of economics at the beginning of the course, also had no impact on the efficacy of the flipped classroom environment. We can identify no differential treatment effect depending on initial economics knowledge.

We have argued above that we do not believe there to be any selection effects across the four sections. Students likely do not know that the professor is engaged in a pedagogical experiment when they register for Principles of Microeconomics during the summer between high school and college, and they do not learn what type of section (flipped or traditional) they are in until the second class, at which point it is quite difficult to change sections.[xxv] But in the fall of 2014, a higher number of students dropped the course after the second class, which is in part why we exclude this data from our baseline analysis.[xxvi]

However, if students truly are assigning themselves randomly across the four sections each year, then the researcher does not need to be concerned about unobserved differences between the two groups of students (flipped versus traditional class format) and does not need to control for characteristics such as ACT score, TUCE scores at the beginning of the semester, adjusted high school GPA, gender, and Pell grants in order to identify a treatment effect. A truly randomized design does not require individual controls, as students with different characteristics are distributed randomly across sections.[xxvii] Table Six A presents our results with such a simplified regression for the three years (twelve sections) of data that are included in our baseline analysis (Table Two). The treatment effect remains statistically insignificant. The estimates continue to suggest that students in noon sections (default section) perform worse than students in other sections. Students in 2017 appear to have performed better than students in other years, which is consistent with an unusually strong entering freshmen class as noted by the University of Richmond Admissions Office. Since our results do not appear to be affected by the absence of additional controls, we repeat the analysis without the added controls for all four years of data (sixteen sections and 333 students), reporting the results in Table Six B. The treatment effect is still statistically insignificant. The strong performance of the 2017 students and the relatively weak outcomes of the noon sections remain. The results of Table Six B are also consistent with Table Two, where more extensive controls are available. The relatively weak performance of the noon sections remain. The positive effect of the 2017 cohort in Table Six B, consistent with a rising admissions profile of students, disappears with the greater controls in Table Two.

[Insert Tables 6A and 6B about here]

DISCUSSION

There are at least two reasons why the coefficient estimates of the treatment effect of the flipped classroom approach might be biased downward. First, the control group already has some elements of a flipped classroom, including seven classroom experiments/group activities. Each section met 41 times in 2015 and 42 times in 2016 and 2017; two classes were dedicated to midterm examinations. Thus, the experiment may be better described as identifying the impact of a completely flipped classroom when compared to a classroom with some elements of active learning already in place.[xxviii] Second, each year there appear to have been one or two students in the control group who had a roommate in the treatment group, so one or two control group students might have also had access to the video lectures. Access to the online lectures was limited to those who could sign in to the Blackboard course for the treatment sections. Unfortunately, Blackboard does not report the amount of time each student spends watching linked videos, which would have been a useful control.

Even with the above caveats, there are several possible interpretations of the above data and results. First, 237/333 students from three/four years of data may not be sufficient to identify accurately any treatment effects. However, given that nearly all of our results do not even approach statistical significance and the point estimates are quite close to zero, additional observations seem unlikely to change this result. Second, positive outcomes of the flipped classroom found in other college and university settings might be a function of the initial class size or the academic setting.[xxix] In other words, flipping a classroom might work well in large classes in which students do not stay connected during a lecture, while at the school studied in this paper the existence of small sections might go a long way toward keeping students engaged even with a mainly traditional lecture format. Third, the flipped classroom approach might be most successful in courses with a large standard deviation of ability or initial knowledge; traditional lectures may be more limited in navigating such differences among students, as the lectures are usually pitched to the students in the middle. Fourth, it may take time for students to adjust their learning and studying approach to get the most out of a flipped classroom environment. It is possible that if the same students had been exposed to the flipped classroom format in another course, the flipped class might have positive effects on learning.

It is also worth noting that slightly different treatment or control classroom designs might lead to different results, since there are not clearly ideal versions of either traditional classroom or flipped classrooms. Perhaps different online videos would improve the learning of students in treatment sections. Additional out of classroom video assignments may have caused some students to spend less time reading the text and thereby score worse on chapter quizzes related to the text. Perhaps the control classrooms already included more active learning than traditional classrooms elsewhere. One can imagine many different combinations of treatment and control classroom designs.

Finally, although the upper range of the 95-percent confidence interval on our point estimates overlaps with the smaller positive impacts found in the literature (see Calimeris 2018 and Wozny, Balser, and Ives 2018), there are several possible reasons for the overall discrepancy between our results and previous studies that find a greater positive impact of the flipped classroom format. The improved learning in a flipped class that other studies find may be due to the additional assignments for the students in the flipped class, which provide a greater incentive for students in those classes to complete work prior to class. It would be useful to test this hypothesis in future research by comparing student outcomes across flipped classes with and without such incentives. Alternately, positive outcomes from a flipped class environment found in other research might be capturing an overall time trend, as instructors improve or focus on many pedagogy elements while implementing a new approach. Although the year effects were generally not statistically significant when controlling for student attributes, the instructor in this experiment has noted rising scores from student evaluations in both the treatment and control groups over the course of this experiment.

It is also possible that, even though they do not appear to learn more in a flipped classroom format, students will maintain their economic knowledge for a longer period when they have learned it via more active study. Such a hypothesis would require additional testing at a later date and is a useful direction for future research on the flipped classroom. Of course, there are significant logistical challenges to such a study, and larger sample sizes and additional controls may be necessary to identify learning effects with so many other additional post course influences.

Although the flipped class format does not appear to have any impact on our primary outcome measures, our analysis shows that it does impact two measures of learning during the semester. The positive effect on the first midterm grade may indicate that the flipped classroom does enhance at least short-term learning. In this case, the lack of effect on the second midterm or the primary outcome measures may be due to students in the traditional classroom remedying deficiencies that the first midterm helped them identify. On the other hand, it may also be the case that students in the flipped classroom had engaged in more problem-solving practice and were therefore more familiar with the way questions were presented on the midterm and the types of answers that were expected. If this is the case, then by the second midterm all students, regardless of class format, would have seen enough problems so that any advantage the students in the flipped class had on the first midterm dissipated. In a smiliar vein, the negative effect of the flipped class on online quiz grades could be due to question types and formats that differ substantially from those worked in the classroom and thus require a different problem-solving approach.

Ultimately, researchers would like to ascertain not only if students are learning more from a particular pedagogical approach, but if students are learning more efficiently. Even if students do learn marginally more in a flipped classroom, which we have not conclusively established, it appears to come at the cost of more time spent on task, given students' self-reported hours spent on the course via both standard student evaluations and supplemental evaluations. For years 2015 and 2016, students in the treatment sections reported spending about two hours more per week on the course (from about 7.5 hours to 9.5 hours). For 2017, students in the treatment group surprisingly reported spending marginally less time on the course. These official student evaluation reports are reasonably consistent with students in our supplemental evaluations reporting about 2.6 hours per week watching online lectures, presuming some subsequent times savings when completing online quizzes and problem sets and studying for exams. Future research that accurately collects data about how long students spend watching the online lectures and completing online work would be useful in fully understanding the tradeoffs of a flipped class.[xxx]

In the end, flipping a course is not an either/or decision. There is perhaps and optimal degree of active learning, in which case comparing a conventional course that includes modest active learning with a flipped course might be considering only two locations on the flipped classroom continuum. The optimal level of active learning in a particular course could be somewhere in the large space between these two locations. Or as Boyle and Goffe (2018) argue, what is most important to improve learning is not active learning by itself or flipping a classroom, but rather specific techniques such as just in time teaching assignments, clicker questions to practice key concepts, increased efforts to connect new ideas to old ones, and more metacognition reflection. The instructor in this paper's experiment certainly enjoys teaching via active learning but in the future may seek a compromise between his traditional lecture-based course, which includes some classroom experiments, and a completely flipped classroom.

NOTES

[i] Indeed, one can imagine that in preparing for a flipped classroom approach, the instructor spends significant time prior to its implementation refining lectures, preparing videos, designing assignments, and generally thinking about how to teach the course well. It may, then, simply be the additional focus on how to teach a class that leads to improved outcomes, rather than the flipped format *per se*.

[ii] Bishop and Verleger (2013) and O'Flaherty and Phillips (2015) provide broad surveys of the flipped classroom in general.

[iii] A flipped-blended class is one in which students watch video lectures outside of class and use class time for problem solving in groups (flipped) but they spend less time in class and more time working online than in a traditional class (blended).

[iv] In addition, students were incentivized to attend class in the flipped but not the traditional format in Lombardini, Lakkala, and Muukkonen (2018).

[v] This conclusion is supported by some of the results in Calimeris and Sauer (2016), which show that when "time spent" variables are included in the estimation, most of the significant differences between flipped and traditional class formats disappear. Furthermore, Lombardini, Lakkala, and Muukkonen (2018) report, based on surveys, that students found the pre-tests more valuable than the lectures in the flipped class.

[vi] A time of day effect could be found, for example, if students in a section taught at noon demonstrate less learning not because of the class format but because they are hungry.

[vii] It is, of course, possible that a student in a traditional section gained access to online lectures if a roommate or teammate was in a treatment section and shared his login information. We are skeptical that such efforts were made to spend time watching online lectures that mirrored the traditional lectures a student in a control section observed.

[viii] The decision to make the noon and 1:30 sections our treatment sections in year one was made for no reason other two sections had to be chosen, and it was easier if one section immediately followed the other. This decision was made before there was any knowledge of the registered students, who are primarily first year students. Part of the IRB procedures do not allow the instructor any access to student academic background information such as SAT scores or high school gpas until semester grades were submitted. The instructor also does not know which students consented until after grades were submitted.

[ix] This is almost certainly true of first-year students, who in general do not have access to information from prior students in the course when they register. It is possible that older students might have known about the experiment prior to registering, but they still would not have known whether the section for which they registered was a treatment or control section. Furthermore, only four percent of students were non-freshman.

[x] Students are required to take Principles of Microeconomics before Principles of Macroeconomics at the University of Richmond.

[xi] The final exam, which included the post-course TUCE, was administered on three or four different days, as determined by the University Registrar. A student's course meeting time gave a default day for the final exam, but students could take the final with another section if they desired. Each exam had the same total number of points available.

We control for the day on which students took the exam. The final exam grade was curved in each year by adding the same number of points in a particular year to all students' exam grade that year.

[xii] T-tests reveal that the difference in means between years was not significant at the 10 percent level or lower.

[xiii] We explicitly include these year dummies instead of using fixed effects in order to identify any potential time trend.

[xiv] The adjustment standardizes the GPA to account for differences in the rigor of courses.

[xv] The first exam was administered on either the 1st or 2nd day of the exam period (on the first day in 2015 and 2017 and on the second day in 2016), while the last exam was administered on day 5 or day 4 or day 8 of the exam period (2015, 2016, and 2017, respectively).

[xvi] The significant drop in the share who took the exam on any given day may be due to the fact that there were 4 exam days available in 2017 rather than 3.

[xvii] Because fractional response models admit only values between 0 and 1, we had to top-code the course grade variable (1 student earned a 1.003 which we coded as a 1) and bottom-code the gap variables (2 students did worse on the TUCE post-test and we entered 0 for their gap values).

[xviii] Even when controlling for year and time of a section, the errors for students in a given

levels of academic preparation. Clustered standard errors are corrected for correlation across errors within a cluster as well as for heteroscedasticity.

[xix] There are also two students in the 2016 sections for whom the adjusted high school GPA variable was not available, and they are also excluded.

[xx] Our R-squared values (which are only calculated for the linear regressions) are broadly consistent with the literature: they are around .3, which is quite close to the value in Balaban, Gilleskie, and Tran (2015) and Calimeris (2018) and in the middle of the range in Olitsky and Cosgrove (2016).

[xxi] Because the gap outcome variable incorporates the TUCE pre-course score, we do not include the pre-course score as a control variable for regressions with gap as the dependent variable.

[xxii] We do not include whether a student took the exam early or late in the exam period here since the final exam took place after all this work was completed.

[xxiii] We might also be concerned that the flipped classroom format might incentivize students to spend more time studying overall so that any increase in performance is the result of spending more time studying rather than the result of a different pedagogical technique. We therefore also include the self-reported time spent studying variables as controls when estimating equation (1) but find that there is no statistically significant impact with the exception of hours spent on the online chapter quizzes, which has a

[xxiv] There are also some students who report spending more than 20 hours per week on the course. This seems improbable due to students' other commitments (both curricular and extra-curricular), casting even more doubt on these self-reported numbers. In order to account for these presumably erroneous estimates, we winsorize the top 5 percent of the time spent studying data.

[xxv] Indeed, when running a probit estimation of the Treatment variable on our student characteristic data, only the gender had any predictive value (see Table A1 in the Appendix). The ACT score, TUCE pre-test score, Pell Grant amount, and high school GPA were all far from conventional levels for statistical significance. This supports our contention that there was no selection into sections and that ability and aptitutde were not correlated with the pedagogical approach. While the gender variable is statistically correlated with being in a treatment section, this is likely due to women registering for 9am classes more often than men since in two of our three baseline years (2015 and 2017), the 9am class was a treatment section. Across all three years, the 9am section had the most women (25, compared to 17 and 20 in the other sections) and also fewer men than two of the other sections (37 compared to 41 and 43 in the noon and 3pm sections).

[xxvi] We are also missing two of the control variables—adjusted high school GPA and Pell grants—for the 2014 data.

[xxvii] We thank participants of the 2016 Conference on Teaching Research in Economic

[xxix] Becker and Proud (2018), for example, find that the impact of a flipped tutorial differs across the two universities where they implement the method and note that this may be due to tutorial size, academic environment, or other factors.

[xxx] Of course, the instructor's start-up time in recording lectures and preparing active learning materials is an additional consideration.

REFERENCES

Abadie, A., Athey, S., Imbens, G.W. and Wooldridge, J., 2017. When should you adjust standard errors for clustering? (No. w24003). National Bureau of Economic Research.

Balaban, Rita A., Gilleskie, D., and U. Tran. 2016. A Quantitative evaluation of the flipped classroom in a large lecture principles of economics course. *Journal of Economic Education* 47 (4): 269-87.

Becker, R. and S. Proud. 2018. Flipping quantitative tutorials. *International Review of Economics Education.* 29: 59-73.

Becker, W. E. and M. Watts (Eds.). 1998. *Teaching economics to undergraduates: Alternatives to chalk and talk*. Edward Elgar, Northampton, MA.

Becker, W. E., Watts, M., and S. R. Becker (Eds.), 2006. *Teaching economics: More alternatives to chalk and talk*. Edward Elgar, Northampton, MA.

Betihavas, V., Bridgman, H., Kornhaber, R. and M. Cross. 2016. The evidence for 'flipping out': A systematic review of the flipped classroom in nursing education. *Nurse education today*. 38: 15-21.

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

Boyle, A. and W. L. Goffe. 2018. Beyond the flipped class: The impact of research-based teaching methods in a macroeconomics principles class" *AEA Papers and Proceedings*. 108: 297-301.

Calimeris, L. 2018. Effect of flipping the principles of microeconomics class: Does scheduling matter? *International Review of Economics Education.* 29: 29-43.

Calimeris, L. and K. M. Sauer. 2015. Flipped out about the flip: All hype or is there hope? *International Review of Economics Education.* 20: 13-28.

Caviglia-Harris, J. 2016. Flipping the undergraduate economics classroom: Using online videos to enhance teaching and learning. *Southern Economic Journal*. 83 (1): 321-331.

Crouch, C. H. and E. Mazur. 2001. Peer instruction: Ten years of experience and results. *American Journal of Physics*. 69 (9): 970-77.

Giannakos, M.N., Krogstie, J. and N. Chrisochoides. 2014. Reviewing the flipped classroom research: Reflections for computer science education. In *Proceedings of the Computer Science Education Research Conference*: 23-29. ACM.

Hoyt, G.M. and K. McGoldrick. (Eds.), 2012. *International Handbook on Teaching and Learning Economics.* Northampton, MA: Edward Elgar.

Jensen, J.L., Kummer, T.A. and P. D. D. M. Godoy. 2015. Improvements from a flipped classroom may simply be the fruits of active learning. *CBE—Life Sciences Education*. *14* (1) ar5.

Lage, M.J., Platt, G.J. and M. Treglia. 2000. Inverting the classroom: A gateway to creating an inclusive learning environment. *The Journal of Economic Education*. 31 (1), 30-43.

Lombardini, C., Lakkala, M., and H. Muukkonen. 2018. The impact of the flipped classroom in a principles of microeconomics course: Evidence from a quasiexperiment with two flipped classroom designs. *International Review of Economics Education.* 29:*14-28*.

O'Flaherty, J. and C. Phillips. 2015. The use of flipped classrooms in higher education: A scoping review. *The Internet and Higher Education*. *2*5*:* 85-95.

Olitsky, N. H. and S. H. Cosgrove. 2016. The better blend? Flipping the principles of microeconomics. *International Review of Economics Education*. 21: 1-11.

Psihountas, D. 2018. Flipped classrooms and finance—Is this a better way to learn? *Journal of Financial Education*. 44 (1): 1-11.

Roach, T. 2014. Student perceptions toward flipped learning: New methods to increase interaction and active learning in economics. *International Review of Economics Education*. 17: 74-84.

Roehling, P.V., Root Luna, L.M., Richie, F.J. and J. J. Shaughnessy. 2017. The benefits, drawbacks, and challenges of using the flipped classroom in an introduction to psychology course. *Teaching of Psychology*. 44 (3): 183-192.

Van Sickle, J.R. 2016. Discrepancies between student perception and achievement of learning outcomes in a flipped classroom. *Journal of the Scholarship of Teaching and Learning*. 16 (2): 29-38.

Vazquez, J. J. and E. P. Chiang. 2015. Flipping out! A case study on how to flip the principles of economics classroom. *International Advances in Economic Research*. 15: 379-90.

Wozny, N., Balser, C., and D. Ives. 2018. Evaluating the flipped classroom: A randomized controlled trial. *The Journal of Economic Education.* 49 (2): 115-129.

APPENDIX

TABLE A1—Probit Estimation of Treatment as a Function of Student Characteristics

	Treatment
Sex (female=1)	0.437***
	(0.156)
ACT score	0.045
	(0.040)
TUCE pre-test	0.028
	(0.028)
Pell Grant amount	0.024
	(0.053)
Adjusted high school GPA	0.130
	(0.365)
N	237
R-Squared	0.030

Notes: Clustered standard errors are in parentheses.

* $p < .10$; ** $p < .05$; *** $p < .01$

TABLE 1: Summary Statistics—2015, 2016, 2017

Variable	***N***	**Mean**	**Standard Deviation**	**Minimum Value**	**Maximum Value**
		Outcome Variables			
Gap	239	0.562	0.180	0	1
Course	239	0.850	0.070	0.650	1.003
Final exam	239	166.320	19.656	89	209
Problem Solving Questions	239	0.710	0.129	0.307	0.986
TUCEpost	239	22.492	3.362	10	30
Daily Quiz	239	81.183	8.982	50	100
Online chapter Quiz	239	87.671	6.428	59.1	97.44
Problem Sets	234	0.848	0.062	0.581	0.995
Midterm I	239	83.814	9.591	46	102
Midterm II	238	83.107	10.388	58.5	103
		Flipped Classroom Variable			
Treatment	239	0.490	0.501	0	1
		Covariates			
Sex (female=1)	239	0.343	0.476	0	1

Variable	*N*	Mean	Standard Deviation	Minimum Value	Maximum Value
ACT score	238	30.324	2.528	20	36
TUCE pre-test	239	12.839	3.945	4	25
Pell Grant amount (thousands)	239	0.506	1.499	0	5.92
Adjusted high school GPA	237	3.560	0.291	2.46	4
English not first language	239	0.075	0.264	0	1
9am section	239	0.259	0.439	0	1
noon section	239	0.243	0.430	0	1
1:30 pm section	239	0.234	0.424	0	1
3:00 pm section	239	0.264	0.442	0	1
First exam date	239	0.205	0.405	0	1
Last exam date	239	0.360	0.481	0	1
2015	239	0.347	0.477	0	1
2016	239	0.301	0.460	0	1
Average weekly hours spent					
Reading textbook	214	3.375	1.889	0	8
Completing MyEconLab assignments	214	2.159	0.832	0.5	4

Variable	*N*	Mean	Standard Deviation	Minimum Value	Maximum Value
Completing optional MyEconLab work	214	0.448	0.627	0	2
General studying	212	1.563	1.290	0	5
Viewing online lectures	239	1.253	1.601	0	5
Total Hours	239	8.521	5.068	0	27

TABLE 2: Baseline Analysis, End-of-Semester Outcomes—2015, 2016, 2017

	Course Grade	Final Exam	Problem Solving Questions	TUCE Post-test	Gap
Treatment	-0.004	-0.761	-0.007	0.092	0.013
	(0.006)	(1.841)	(0.012)	(0.384)	(0.026)
Sex (female=1)	0.005	-1.897	0.013	-0.593	-0.035
	(0.008)	(2.581)	(0.018)	(0.422)	(0.028)
ACT score	0.007***	2.216***	0.012***	0.263***	0.010*
	(0.001)	(0.297)	(0.002)	(0.082)	(0.006)
TUCE pre-test	0.005***	1.545***	0.008***	0.321***	
	(0.001)	(0.266)	(0.002)	(0.068)	
Pell Grant amount	0.006***	1.384**	0.006	0.273**	0.015**
(in thousands)	(0.002)	(0.583)	(0.004)	(0.112)	(0.006)
Adjusted high school GPA	0.044***	7.002	0.030	0.010	0.012
	(0.015)	(5.444)	(0.033)	(0.649)	(0.033)
9:00 am section	0.013	3.625	0.035**	-0.038	-0.019
	(0.008)	(2.428)	(0.016)	(0.591)	(0.041)

	Course Grade	Final Exam	Problem Solving Questions	TUCE Post-test	Gap
1:30 pm section	0.015**	4.381	0.026	0.342	0.022
	(0.008)	(2.563)	(0.018)	(0.577)	(0.043)
3:00 pm section	0.006	4.731*	0.046***	-0.011	-0.006
	(0.008)	(2.437)	(0.017)	(0.549)	(0.038)
First exam	0.011	3.426	0.032**	0.381	0.018
	(0.010)	(2.147)	(0.014)	(0.489)	(0.035)
Last exam	-0.012	-3.777	-0.014	-0.541	-0.020
	(0.012)	(3.273)	(0.018)	(0.633)	(0.032)
2015	-0.007	0.551	0.018*	-0.752*	-0.037
	(0.007)	(1.911)	(0.010)	(0.410)	(0.030)
2016	-0.002	2.264	0.010	-0.450	-0.030
	(0.008)	(1.845)	(0.012)	(0.354)	(0.022)
N	237	237	237	237	237
R-Squared		0.332		0.304	

Notes: Clustered standard errors are in parentheses.

* $p < .10$; ** $p < .05$; *** $p < .01$

TABLE 3: Mid-Semester Outcomes—2015, 2016, 2017

	Daily Quiz	Online Chapter Quizzes	Problem Sets	Midterm I	Midterm II
Treatment	-0.819	-0.709*	-0.001	2.492***	-0.427
	(1.373)	(0.359)	(0.006)	(0.752)	(1.293)
Sex (female=1)	1.395	0.825	0.005	0.295	0.937
	(0.925)	(0.462)	(0.008)	(1.047)	(1.292)
ACT score	0.388**	0.489***	0.008***	1.399***	1.297***
	(0.167)	(0.120)	(0.001)	(0.219)	(0.183)
TUCE pre-test	0.375**	0.321**	0.006***	0.601***	0.767***
	(0.138)	(0.110)	(0.001)	(0.098)	(0.088)
Pell Grant amount	0.394	0.397*	0.005***	0.220	0.889**
(in thousands)	(0.333)	(0.220)	(0.002)	(0.256)	(0.298)
Adjusted high school GPA	7.707**	3.900**	0.045***	6.064***	8.693***
	(2.515)	(1.770)	(0.014)	(1.609)	(1.988)
9:00 am section	1.587	-0.464	0.017*	1.087	2.864
	(1.700)	(0.680)	(0.009)	(1.092)	(1.839)
1:30 pm section	1.193	0.074	0.015*	1.237*	2.806

	Daily Quiz	Online Chapter Quizzes	Problem Sets	Midterm I	Midterm II
	(1.683)	(0.537)	(0.008)	(0.672)	(1.669)
3:00 pm section	-0.204	-1.218**	0.004	0.245	0.348
	(1.593)	(0.473)	(0.008)	(1.055)	(1.670)
2015	-3.708***	-2.177***	-0.009*	-0.477	-0.840
	(0.664)	(0.442)	(0.004)	(0.865)	(0.941)
2016	-4.724***	-1.809***	-0.005	-0.768	2.395
	(1.426)	(0.404)	(0.007)	(0.809)	(1.364)
N	237	237	237	237	236
R-Squared	0.203	0.183		0.354	0.373

Notes: Clustered standard errors are in parentheses.

* $p < .10$; ** $p < .05$; *** $p < .01$

TABLE 4: Time Spent Studying—2015, 2016, 2017

	Hours per Week Reading Textbook	Hours per Week Taking Chapter Quizzes	Hours per Week On Optional Online Exercises	Hours per Week on General Studying	Total Hours per Week
Treatment	0.013	-0.044	-0.140	0.036	3.102***
	(0.207)	(0.061)	(0.079)	(0.266)	(0.855)
Sex (female=1)	0.198	0.121	0.022	0.228*	0.851
	(0.240)	(0.115)	(0.085)	(0.119)	(0.675)
ACT score	-0.087	-0.021	-0.047***	-0.120**	-0.226
	(0.067)	(0.021)	(0.014)	(0.039)	(0.151)
TUCE pre-test	-0.060	-0.051***	-0.014	-0.017	-0.150
	(0.045)	(0.011)	(0.009)	(0.025)	(0.100)
Pell Grant amount	-0.110	-0.091***	-0.018	-0.095**	-0.517**
	(0.112)	(0.018)	(0.031)	(0.035)	(0.234)
Adjusted high school GPA	0.303	-0.088	0.132	0.235	0.653
	(0.349)	(0.174)	(0.123)	(0.225)	(0.982)
9:00 am section	-0.365	-0.408**	-0.144	-0.600*	-1.744*
	(0.274)	(0.141)	(0.156)	(0.274)	(0.884)

	Hours per Week Reading Textbook	Hours per Week Taking Chapter Quizzes	Hours per Week On Optional Online Exercises	Hours per Week on General Studying	Total Hours per Week
1:30 pm section	-0.214	-0.055	-0.077	0.218	-0.126
	(0.152)	(0.115)	(0.168)	(0.311)	(0.710)
3:00 pm section	-0.847***	-0.152	-0.118	-0.096	-2.004*
	(0.238)	(0.104)	(0.143)	(0.304)	(0.971)
2015	-0.438***	-0.054	0.189	-0.081	-0.707
	(0.136)	(0.108)	(0.144)	(0.160)	(0.704)
2016	0.029	-0.046	0.257**	-0.209	-0.113
	(0.187)	(0.089)	(0.092)	(0.253)	(0.958)
English not first language	1.284**	0.924***	0.352*	0.954***	3.336**
	(0.451)	(0.286)	(0.187)	(0.287)	(1.410)
N	213	213	213	211	237
R-Squared	0.105	0.202	0.153	0.161	0.157

Notes: Clustered standard errors are in parentheses.

* $p < .10$; ** $p < .05$; *** $p < .01$

TABLE 5: Interaction Terms—2015, 2016, 2017

	Course Grade	Final Exam	Problem Solving Questions	TUCE Post-test	Gap
		Early Class Time (9 am)			
Treatment	0.000	-0.290	-0.005	0.253	0.022
	(0.005)	(1.766)	(0.012)	(0.376)	(0.024)
Early Class Time	0.023***	4.901*	0.040**	0.399	0.005
	(0.008)	(2.497)	(0.017)	(0.677)	(0.044)
Treatment x Early	-0.015***	-1.961**	-0.008	-0.671**	-0.037
	(0.006)	(0.856)	(0.008)	(0.295)	(0.025)
N	237	237	237	237	237
R-Squared		0.332		0.305	
		Late Class Time (3 pm)			
Treatment	-0.008	-1.271	-0.009	-0.083	0.004
	(0.006)	(1.972)	(0.012)	(0.424)	(0.030)
Late Class Time	-0.003	3.626	0.041**	-0.390	-0.027
	(0.006)	(2.348)	(0.017)	(0.517)	(0.037)

	Course Grade	Final Exam	Problem Solving Questions	TUCE Post-test	Gap
Treatment x Late	0.015***	1.961**	0.008	0.671**	0.037
	(0.006)	(0.856)	(0.008)	(0.295)	(0.025)
N	237	237	237	237	237
R-Squared		0.332		0.305	
		Female			
Treatment	-0.005	-0.723	-0.005	0.271	0.034
	(0.008)	(2.515)	(0.018)	(0.508)	(0.037)
Female	0.002	-1.837	0.016	-0.312	-0.003
	(0.011)	(3.640)	(0.025)	(0.649)	(0.036)
Treatment x Female	0.005	-0.111	-0.005	-0.520	-0.059
	(0.015)	(4.847)	(0.039)	(0.776)	(0.051)
N	237	237	237	237	237
R-Squared		0.332		0.305	
		Pell Grant			
Treatment	-0.004	-0.585	-0.010	0.257	0.024
	(0.006)	(1.942)	(0.013)	(0.424)	(0.029)

	Course Grade	Final Exam	Problem Solving Questions	TUCE Post-test	Gap
Pell Grant	0.006*	1.561	0.003	0.438***	0.026***
	(0.003)	(1.039)	(0.006)	(0.125)	(0.007)
Treatment x Pell	0.001	-0.359	0.007	-0.337	-0.022*
	(0.004)	(1.225)	(0.008)	(0.193)	(0.012)
No	237	237	237	237	237
R-Squared		0.332		0.309	
		Low High School GPA			
Treatment	0.002	1.105	-0.010	0.183	0.014
	(0.010)	(3.271)	(0.025)	(0.516)	(0.031)
High School GPA	0.038**	4.688	0.034	-0.103	0.011
	(0.016)	(5.914)	(0.037)	(0.769)	(0.042)
Treatment x Low GPA	-0.010	-3.531	0.006	-0.173	-0.002
	(0.011)	(4.363)	(0.031)	(0.533)	(0.031)
N	237	237	237	237	237
R-Squared		0.334		0.304	
		Low ACT			

	Course Grade	Final Exam	Problem Solving Questions	TUCE Post-test	Gap
Treatment	0.000	0.878	-0.003	0.805	0.070
	(0.009)	(2.421)	(0.017)	(0.672)	(0.047)
ACT	0.007***	2.038***	0.012***	0.186	0.004
	(0.001)	(0.462)	(0.003)	(0.112)	(0.008)
Treament x Low ACT	-0.005	-2.427	-0.006	-1.056	-0.083*
	(0.012)	(2.864)	(0.017)	(0.722)	(0.050)
N	237	237	237	237	237
R-Squared		0.333		0.312	
		Low TUCE Pre-test			
Treatment	-0.010	-3.000	-0.027	-0.032	-0.005
	(0.009)	(2.375)	(0.017)	(0.464)	(0.029)
TUCE pre-test	0.006***	1.785***	0.010***	0.335***	
	(0.001)	(0.389)	(0.002)	(0.092)	
Treatment x low TUCE	0.012	4.889	0.039*	0.270	0.039
	(0.011)	(3.484)	(0.020)	(0.788)	(0.028)
N	237	237	237	237	237

	Course Grade	Final Exam	Problem Solving Questions	TUCE Post-test	Gap
R-Squared		0.337		0.305	

Notes: Clustered standard errors are in parentheses.

* $p < .10$; ** $p < .05$; *** $p < .01$

TABLE 6A: Section Time and Year Controls Only—All Years

	Course Grade	Final Exam	Problem Solving Questions	TUCE Post-test	Gap
Treatment	0.003	-0.114	0.004	0.154	0.009
	(0.007)	(2.375)	(0.016)	(0.430)	(0.023)
9:00 am section	0.037***	10.795***	0.071***	0.800	-0.003
	(0.009)	(2.955)	(0.020)	(0.555)	(0.033)
1:30 pm section	0.022***	6.222**	0.042**	0.369	0.024
	(0.008)	(2.627)	(0.020)	(0.500)	(0.036)
3:00 pm section	0.010	6.776**	0.056***	0.188	0.002
	(0.009)	(2.847)	(0.019)	(0.551)	(0.031)
2015	-0.017***	-1.528	0.005	-1.072***	-0.047**
	(0.005)	(1.481)	(0.010)	(0.327)	(0.022)
2016	-0.017**	-1.319	-0.006	-0.943*	-0.053**
	(0.008)	(2.462)	(0.017)	(0.455)	(0.022)
N	239	239	239	239	239
R-Squared		0.04		0.031	

Notes: Clustered standard errors are in parentheses.

* $p < .10$; ** $p < .05$; *** $p < .01$

TABLE 6B: Section Time and Year Controls Only—2015, 2016, 2017

	Course Grade	**Final Exam**	**Problem Solving Questions**	**TUCE Post-test**	**Gap**
Treatment	0.007	2.196	0.018	0.383	0.012
	(0.006)	(2.093)	(0.015)	(0.326)	(0.015)
9:00 am section	0.024**	6.087	0.036	0.428	-0.009
	(0.010)	(3.578)	(0.025)	(0.510)	(0.024)
1:30 pm section	0.019***	5.583**	0.022	0.620	0.025
	(0.007)	(2.231)	(0.020)	(0.423)	(0.027)
3:00 pm section	0.012	6.395**	0.043**	0.367	0.005
	(0.008)	(2.589)	(0.020)	(0.430)	(0.022)
2015	0.014*	2.409	-0.044**	0.330	0.022
	(0.008)	(2.885)	(0.020)	(0.467)	(0.021)
2016	0.014	2.686	-0.054*	0.453	0.016
	(0.012)	(4.126)	(0.028)	(0.583)	(0.021)
2017	0.031***	3.879	-0.049**	1.383***	0.069***
	(0.008)	(2.668)	(0.019)	(0.376)	(0.009)

			Problem Solving		
	Course Grade	Final Exam	Questions	TUCE Post-test	Gap
N	317	317	317	317	317
R-Squared		0.024		0.028	

Notes: Clustered standard errors are in parentheses.

* $p < .10$; ** $p < .05$; *** $p < .01$