

2-12-2015

Flipping the Classroom and Student Performance in Advanced Statistics: Evidence from a Quasi- Experiment

Michael Touchton
Boise State University

Flipping the Classroom and Student Performance in Advanced Statistics: Evidence from a Quasi-Experiment

Michael Touchton

Department of Political Science
Boise State University

Abstract

I administer a quasi-experiment using undergraduate Political Science majors in statistics classes to evaluate whether “flipping the classroom” (the treatment) alters students’ applied problem solving performance and satisfaction relative to students in a traditional classroom environment (the control). I also assess whether general student characteristics such as when/where students took the prerequisite course, GPA and gender influence performance. I find flipping the classroom gives students statistically significant advantages in difficult, applied areas emphasized in class. Furthermore, students in the flipped classroom feel they learned more and enjoyed the course more than those in a traditional classroom. I argue students’ affective preference for a flipped classroom is important for student motivation, recollection and future use of quantitative data analysis. Flipping the classroom entails high start-up costs, but it can merit implementing to improve both effective and affective instructional outcomes.

Keywords: flipping the classroom, statistics education, quasi-experiment, active learning

The Teaching and Learning Question

Students in my advanced statistics classes have the most difficulty grasping the most important concepts. These students almost always perform well at lower-order cognitive tasks such as memorizing a definition or calculating a numerical value, but considerably less well when asked to apply the material in context. This troubles me as an instructor because students acquire useful skills through application- not memorization and regurgitation. Yet, it is clear higher-order cognitive tasks surrounding applications of the material are difficult not just for Political Science students in statistics classes, where the material may be relatively foreign, but for students in general. Harnessing theoretical concepts and applying them in practice is one of the greatest conceptual leaps students make in the course of their education and is therefore one of students’ greatest challenges. It is also one of the most important learning outcomes because the point of education is not to recite facts and figures, but to apply concepts and skills to solve problems. Students can always look up information. However, learning how to use information purposefully will help students throughout their lives. It is therefore critical to understand the factors that influence students’ ability to move between theory and practice, from lower-order cognition to higher-order cognition. My research evaluates course delivery strategies designed to bridge the gap between lower-order and higher-order cognitive learning outcomes. Specifically, I use a quasi-experiment in two sections of advanced statistics classes to determine if flipping the classroom, a popular instructional technique in K-12 and some university-level educational environments, influences student performance on applied statistics assignments relative to traditional instructional methods.

Flipping the classroom has gained prominence as a course delivery strategy in recent decades. The instructor “flips” the classroom by first asking students to perform tasks traditionally limited to the classroom, such as listening to lectures and watching presentations, at home. The classroom flip is completed when students perform traditional homework, lab and research paper activities under the instructor’s supervision in class- often with help from their peers. This allows the instructor to intervene at the point of need, when questions about the assignments arise, and to structure learning through a goal-oriented, applied framework with frequent feedback. The results of my research demonstrate several ways flipping the classroom gives students a statistically significant performance advantage over students learning in a traditional classroom. Additionally, I find students in the flipped classroom expressed greater satisfaction with their own learning outcomes and the course in general than their peers in a traditional

classroom. I argue students' affective preference for a flipped classroom is important for student motivation, recollection and future use of quantitative data analysis; especially among Political Science undergraduates who may not consider these techniques attractive to learn or useful to deploy. Flipping the classroom can therefore be worth implementing for two important reasons: immediate effect and long-term affect.

The remainder of the paper proceeds as follows: First, I present arguments surrounding flipping the classroom and describe how a flipped classroom works in practice. Then I describe the study I designed to assess whether the technique influenced my students' outcomes. Finally, I discuss the results of analysis, identify lessons for teaching and offer suggestions for transferring those lessons to other contexts.

Why Flip the Classroom?

Education literature from the 1990s to the present suggests a flipped classroom and other pedagogies that support active learning may benefit students more than a traditional lecture-in-class, homework-at-home model (King, 1993; Ronchetti, 2010; Berrett, 2012; Bergman and Sams, 2012; Brunzell and Horejsi, 2011; Fulton, 2012). This is because students in a flipped class perform the lower-order, easier tasks from Bloom's revised taxonomy (Anderson et al., 2005) outside of class and the higher-order, more difficult tasks in class, with instructor and peer support. For example, reading course material and listening to lectures are both relatively passive activities that do not require as much thought as many homework or lab assignments because students rarely do anything while reading or listening to a lecture other than to read or listen. Students' understanding is only tested (and improved) when they apply the course material to a particular context in an activity like a homework problem set. Activities require much more higher-order cognition than reading or listening to a lecture. Active learning is more difficult than passive learning, but the payoff is potentially greater because activities cement concepts in students' minds more permanently than if students only read the material. In a flipped classroom students read, listen to pre-recorded lectures and/or watch presentations on their own time. Then students apply this material to higher-order cognitive learning activities in class, which proponents of a flipped classroom argue improves learning outcomes.

The shift in emphasis a flipped classroom requires is not new: it only represents an additional step in the direction of higher-order cognition that is already underway. For instance, the shift from reading in class to lecturing in class in order to supplement and explain the reading occurred centuries ago in many countries. Of course, this shift reflects the belief reading foundational course material during class is an inefficient use of instructors' time. Instead, students now read the material on their own and attend class to hear an instructor's explanation or participate in an instructor-led discussion. Flipped classrooms take this shift a step farther and devote class time to applying the material: the point at which learning truly occurs.

Proponents argue flipping the classroom improves learning efficiency and produces outcomes superior to traditional instructional techniques. For example, Hake (1998), Crouch and Mazur (2001), Mazur (2009) and DesLauriers et al. (2011) show how some aspects of the flipped-classroom such as learning before lecture, peer instruction and active-learning improve students' Physics test scores. Moravec et al. (2010) provides similar evidence for an introductory Biology class. These studies all occur in the natural sciences, with the exception of Lage et al. (2000) in Economics, and all use large introductory classes as research laboratories. My study is thus the first to formally present the results of this technique for an upper-division statistics course and, as far as I know, for a Political Science class.

Flipping the classroom is theoretically attractive to educators across disciplinary boundaries and among levels within disciplines because it creates an environment emphasizing goal-directed practice and feedback, which Ambrose et al. (2010) tie directly to improved learning outcomes. In this case the study-goal is clear: students learn material to use it in classroom learning activities. This is not to say all students actually watch slideshows and lectures before class, but they all have the opportunity to do so, just as with assigned reading in a traditional classroom. Theoretically, students who are more familiar with the material will likely perform better on their assignments than those that lack such familiarity. Next, students practice their skills under an instructor's and/or peers' supervision with the goal of mastery prior to further application on exam problems or research projects. Most instructors already have an idea of where problems with these assignments commonly occur, but having students read/watch the lecture material at home gives instructors enough free time to pause class and address students' immediate needs. The result is improved learning outcomes because instruction is most effective when delivered precisely as students are thinking about and working toward their goal, rather than after they have moved on to new material. Finally, the flipped classroom can stimulate peer-instruction among students. Formal working groups

and/or informal exchanges let students maintain their progress when the instructor is busy providing individualized instruction elsewhere. The flipped classroom can thus give students the opportunity to teach one another, which requires them to learn the material more thoroughly than a simple homework application would otherwise demand. In this way, the flipped classroom can also cultivate a broader sense of teamwork among students and render it easier for them to form study groups outside of class.

Goal-directed practice and immediate feedback are especially helpful to students in research methods classes, such as statistics, for several reasons. First, statistics classes impart a skill and require students to master a set of tools rather than a subject's theoretical intricacies. Statistics instruction thus lends itself to activities where students work backwards from a correct, numerical answer to identify mistakes in calculation, practice interpreting very similar calculations using different data and learn routines using software to analyze data. Many statistics classes also encourage deep thinking about the theories behind inferential statistics, but goal directed practices and immediate feedback are especially important for the mechanical and interpretive aspect of the skills students acquire in statistics classes. Performance on these activities improves rapidly with practice, which a flipped classroom provides. Students can learn to use the appropriate formula, the appropriate software command or the appropriate interpretation of statistical output in a matter of minutes, not hours. As a result, immediate feedback on homework in a flipped statistics classroom can potentially break bad habits and replace them with good ones, or prevent students from acquiring bad habits in the first place. In contrast, timely feedback may be difficult to provide in traditional classrooms and may require significant time to incorporate in other, subject-bound classes.

In addition to performance outcomes, students may enjoy a flipped classroom more than a traditional one. Students are likely to feel more comfortable with the course material in a flipped classroom as they benefit from immediate help with a course's hardest tasks. Comfortable students exhibit affective responses in favor of classes where they are comfortable and can then dispense with fear as a barrier to education, especially in their most difficult classes (Ambrose et. al., 2010). The result, theoretically, is not only a greater dispensation for students to learn more in these specific flipped classes relative to traditional classes, but also to revisit the material and augment their education in the future simply because they enjoyed the course more. Affective responses to educational techniques can therefore be important factors that motivate instructors' decisions to flip the classroom.

Methods: The Flipped Classroom as Treatment

I used two sections of advanced statistics classes at Boise State University to assess the extent to which a classroom flip would work in my course- one that challenges students to leave their comfort zones as Political Science majors and engage the higher levels of cognition in Bloom's taxonomy in new ways.¹ More importantly, I use this research design to reveal information about flipping the classroom in general. Prior to my research I had no clear predisposition in favor of or against flipping the classroom. As discussed above, I see potential advantages and disadvantages to the instructional model. However, I felt the arguments in support of a flipped classroom were strong enough to merit testing whether the technique improved students' performance and warranted future use in my statistics courses.

Our department requires majors to take introductory and advanced statistics courses. First, the introductory class teaches students how to create hypotheses regarding political relationships and how to use several basic statistical techniques to test those hypotheses. The techniques we teach include the following: measures of central tendency and dispersion, cross-tabulations, difference of means tests, analysis of variance, linear correlation and bivariate regression using OLS. Finally, students learn to work with the SPSS software platform and use it to perform limited data analysis. The second course in the sequence begins with bivariate regression and progresses through multivariate regression with OLS, diagnostics for the violation of OLS assumptions, remedies for violated assumptions and logistic regression. The second course improves student proficiency in SPSS and also requires an article-length research paper where students gather data, test hypotheses, analyze results, etc. surrounding a research question of their choice.

¹ Boise State University is a public, metropolitan research university in the U.S, with approximately 23,000 students. Many of our students are first-generation college attendees and many (45%) are of non-traditional age.

Students in one of my fall, 2012 Advanced Statistics sections experienced the course through a flipped classroom format: these students read all textbook material at home (like normal), but also watched mini-lectures and slideshows I normally give in class on their own time. This material demanded approximately an extra hour and a half of work for students before attending class- assuming the students only watched or listened to the material once. These students then completed their homework problem sets and lab exercises in class throughout the semester. The homework and lab assignments were tied directly to the readings, with my lectures and slideshows designed to contextualize elements from the textbook through examples and to simplify and explain difficult concepts.² Having the students explore this material at home gave me the freedom to offer mini-lectures addressing common problems in class as well as to assist individual students at the point of need- when they were stuck on a homework or lab problem. Flipping the classroom also gave me the time to facilitate higher-order learning surrounding the homework and lab assignments by leading extensive conceptual discussions of the problem sets. The goal here was to give students a deep understanding of what homework and lab assignments represented beyond simple arithmetic exercises and encourage learning as a result. I intended the homework and lab assignments for individual completion, but I allowed and encouraged informal peer-instruction during class when questions arose and I was providing assistance to other students.

In addition to regular homework and lab assignments, I devoted the final portion of the semester to students' research projects, wherein each student designed an investigation of an outstanding Political Science question. These projects required a literature review, hypotheses development, quantitative data collection and multivariate regression model specification for the phenomenon in question. Then students evaluated their models' performance and applied the results to policymaking. For me, this assignment was the most important of the course because students applied statistical techniques they acquired in class to a real-world problem- all following the model of a professional, Political Science journal article. The assignment thus served as a foundation for future application of quantitative analytic skills and also helped me evaluate students' performance because it covered a full range of quantitative analytic skills including operationalizing variables, selecting statistical tests to evaluate hypotheses, assessing model performance, evaluating regression assumptions and critically applying the results.

Students in the treatment (flipped) classroom listened to or watched lecture material on different aspects of the research paper at home and then worked on their papers during class time, with the opportunity for me to give mini-lectures and for students to ask questions directly. In contrast, students in the control (traditional) section of the Advanced Statistics class read their textbooks at home and attended lecture where I delivered the same mini-lectures I had recorded for the flipped classroom and presented the same Power Point slideshows. Then students completed their homework and lab exercises at home, as in the previous semesters I had taught the course. Additionally, these students watched me deliver lectures and present material on their research papers in person. They then worked on their research papers outside of class until the due date at the end of the semester.

Shifting from a traditional to a flipped classroom carries considerable start-up costs. The front-end workload for reconfiguring one's course delivery is indeed high and prohibits the shift from a traditional classroom in many cases. For example, one must either acquire the technical skills necessary to capture lectures or record voice-overs in slideshows or work with a support staff that can assist in this capacity. Many universities now provide support for creating rich media course material, but there are also many user-friendly, inexpensive software products available that make DIY media production possible. I used Explain Everything and Camtasia Relay to translate my traditional course material into a flipped-classroom format but there are many other programs most instructors already have that allow for some form of lecture capture. For instance, Microsoft Power Point allows for voice-overs on slideshows. TAs may be useful from a technical perspective in this area, but most instructors will probably want to give the lectures themselves, which definitely entails a high short-term cost of writing scripts, perfecting one's delivery and posting the material in a way that complies with university access requirements (e.g. lecture captions for hearing-impaired students). These one-time costs are almost unavoidable, but they represent long-term investments: the payoff is large, timesaving benefits as one uses this material in more sections of the class. For me, class preparation time fell to almost zero by the third iteration of the course. The break-even point will differ from instructor to instructor, but I argue the investment might not make sense for a one-off course. However, flipping the classroom can save the instructor considerable time in the long-run if the course is one that she or he teaches regularly.

² I used Philip Pollock III's *The Essentials of Political Analysis* along with his *An SPSS Companion to Political Analysis* in both courses.

Class size and TA support are also directly relevant for making a smooth transition to a flipped classroom. For example, successfully flipping the classroom can be difficult without TA support in large sections. My courses are 75 minutes long, which gives me approximately five minutes per row of seven students and a few minutes for each remaining student who requires additional help. TAs who could also cover the learning activities in class would be extremely helpful in this environment. Yet, I did not have them and I do not consider them a necessary condition for the approach to work so long as sections remain under 50 students. Student heterogeneity in terms of prior math experience is also a concern in the flipped classroom. For example, my students vary widely in their mastery of the introductory statistics material. Yet, completing homework in class through informal groups allows low performers to catch up to median performance levels in the first few weeks of the course.

Ultimately, my paired class research design allowed me to test whether the potential advantages of a flipped classroom described above translated into systematically better student performance and greater enjoyment of the class relative to a traditional classroom. I held constant some elements of instruction such as the subject matter, year in which the course was taught (2012), the instructor and the composition of students (all are Political Science majors taking a required, upper-division course). These students did not know they were subjects in an experiment on instructional techniques or whether they were in the treatment or control groups while the classes were in session or grades were still unreported.³ 83 of 92 students then gave informed consent for me to use their data on classroom performance and their presence in the treatment or control group.

Students also gave me permission to collect data through department records and the registrar's office for other areas that might influence student outcomes to use as control variables in my statistical models. Students' GPA, when and where students took the prerequisite and gender may all be relevant for student outcomes. I accounted for these factors to identify the general determinants of student performance in a small, albeit controlled sample of students in my classes.

My statistical analysis is somewhat limited by my small number of observations of student performance. 83 students released their data for this study out of a total of 92 students. I cannot know whether this sample of students is representative of the broader population of our majors, Political Science majors at other universities, or university students in general. However, I taught both treatment and control sections of the course on the same days with the same amount of time scheduled for each section and no student awareness of whether they were in the treatment or control group. There is no reason to believe the treatment class had a systematically different population than the control class and no reason to believe these samples were systematically different from the broader population of Political Science majors at Boise State University. The only major opportunity to bias the results in this study came from me, the instructor. As stated above, I had no conscious predisposition toward a flipped or traditional classroom when the project began and therefore believe my findings do not stem from any personal bias toward one instructional technique or another.

Modeling Student Outcomes

I modeled student outcomes using propensity-score matching with Ordinary Least Squares (OLS) and robust standard errors to assess whether flipping the classroom produced the desired outcomes of improved performance on applied statistics problems and increased enjoyment of the class.

Dependent Variable 1: Performance on Applied Statistics Research Paper

I created a variable measuring students' performance on their semester-long, applied research project. The variable features a total performance scale ranging from 0 to 100, where 0 indicates the lowest performance and 100 the highest. I evaluated performance primarily through how well students developed hypotheses, tested them against available data, critically assessed the results of estimation and described the implications for policymaking. The mean score for the projects is 81.2 and the standard deviation is 3.58.

³ Boise State University's Institutional Review Board approved this study (Protocol # 025-SB13-010).

I also disaggregated the total research project scores based on the individual components of each paper. Specifically, I used data from my grading rubric to code performance in five areas, each scored 1-20 and worth 20% of the overall project grade⁴: 1) Arguments for the relevance of the central question, the literature review and hypotheses generation. Mean: 16.41, SD: 2.15. 2) Selection and presentation of variables along with description of the research methodology. Mean: 15.48, SD: 2.29. 3) Presentation of the results of estimation using OLS and regression diagnostics. Mean: 13.04, SD: 3.91. 4) Interpretation of the results and a discussion of their implications for theory and policy. Mean: 14.26, SD: 2.72. 5) Clarity and structure of the paper. Mean: 15.68, SD: 1.98.

I exploited variation in these sub-categories to assess the extent to which a flipped classroom influences performance not just in terms of the overall score on an assignment, but in the specific areas to which I devoted more instructional energy and class time as well as those that are traditionally more challenging.

Dependent Variable 2: Teaching Evaluations in Advanced Statistics

I want to know the extent to which students in each section of the course enjoyed their experience. Our majors wear successful completion of the statistics sequence as a badge of honor, but many also view it as a struggle to be overcome rather than as an opportunity to acquire useful skills. I believe students will be more inclined to learn about quantitative analytic methods after they graduate and more inclined to apply these techniques in the workplace if they remember their undergraduate training fondly- not as a struggle, but as an enjoyable, helpful experience. Student enjoyment of a class certainly has a connection to their performance in the class (Ambrose et al., 2010; Baldwin et al., 1997; Battistich, et al., 1995) and striving to improve student performance is my foremost immediate goal in the classroom. I therefore collected anonymous student evaluations of the class for both sections and use the variation between them to assess the role of flipping the classroom for students' affective instructional preferences. These evaluations occurred online for both sections during the last week of classes at the end of the semester before final grades were reported. I encouraged students to fill out these evaluations during class to generate a high response rate, which I achieved: 92% of students responded in the flipped classroom and 87% in the traditional classroom. The scores are out of 5 (5 is highest) and reflect mean student answers to a variety of standard questions about the course and the instructor and their relative value compared to other courses/instructors at the same institution. I report the total mean score, the mean scores for the instructor, for the course and for one constituent question asking students to assess how much they learned in the course being evaluated relative to other courses at the university. These scores appear in Table II, below.

I hypothesized students would enjoy a flipped classroom more than a traditional one for reasons discussed above. However, I considered the possibility a flipped classroom would not conform to their educational expectations as described in Berrett (2012). For example, some students may feel they could replicate the flipped classroom by watching online videos and meeting in groups completely outside of the university environment and as a result feel the professor in a flipped classroom is not adding educational value. Though I took great care to explain why I flipped the classroom, some students may still not have agreed with my argument and might therefore have expressed their preferences for a traditional classroom environment on their course evaluations.

Primary Independent Variable: Flipped vs. Traditional Classroom

I assessed whether a flipped classroom generated distinct student performance outcomes relative to a traditional classroom as discussed above. Students in the flipped (treatment) classroom received a score of 1, whereas students in the traditional (control) classroom received a score of 0 on the variable. I hypothesized students in the flipped classroom would outperform students in the traditional classroom on the applied statistics project.

Control Variables: Elapsed Time between Introductory and Advanced Courses

The length of time between taking the prerequisite and the subsequent course in any course sequence is likely to influence student performance in the subsequent class, all else equal. This seems particularly likely for classes students deem challenging, such as statistics courses within a Political Science major. I collected data on the time, in

⁴ The project description and grading rubric are available in the Technical Appendix

semesters, between completion of the introductory and advanced statistics courses and included this variable in models of student performance. The minimum number of semesters is 0 and the maximum is 14. The mean is 2.7 and the standard deviation is 1.3.

Location Where the Students Took the Introductory Course

Boise State students are not required to take the prerequisite, introductory statistics class in the Political Science Department in order to enroll in the advanced course, but may fulfill the prerequisite through courses in several other departments (e.g. Psychology, Business, Math, etc.). These courses cover broadly similar introductory statistics topics to the classes taught within the Political Science Department. However, I suspect students who take the prerequisite course outside of the department perform worse, on average, than students who take the prerequisite within the department for many reasons. I created a dummy variable to denote whether students took the introductory class outside of our department (0) or within it (1) to test this hypothesis. In my sample 19 students fulfilled the prerequisite outside of the department whereas 64 students took the course within it.

GPA

The conventional wisdom is students with stronger performance in earlier classes, as measured by higher GPAs, will perform better than other students in future classes because, put simply, they are better students. Students with higher GPAs may have better study skills, better time-management skills, more motivation and greater intelligence than students with lower GPAs, all things being equal. At a minimum, any model of student performance should consider previous student performance as a potential determinant of current performance and an appropriate control variable. I therefore collected data on students' GPA (on a 0.0 to 4.0 scale, where 0 is low and 4 is high) and included this variable in my statistical models. The mean is 3.08 and the standard deviation is 0.41.

Gender

Gender may influence student performance in advanced statistics classes in several ways. First, there is evidence male students enroll in STEM disciplines, with heavy emphasis on mathematics, in greater numbers than female students (Sahin et al., 2012 among many others). The reasons why are numerous and controversial. However, the result is the common perception male students are more favorably disposed toward mathematics than female students and will therefore perform better on mathematical tasks. Yet, there is increasing evidence female students outperform male students in high school and university classes in general (Conger and Long, 2010). Applied to students in advanced statistics classes, I controlled for students' gender, in case it is tied to performance by creating a variable measuring students' gender and including it in my models. I coded male students "0" and female students "1". My sample contains 47 males and 36 females.

Estimation Strategy

I estimated relationships between instructional design, student characteristics and performance on an applied statistics project using propensity-score matching with OLS and robust standard errors. Ho et al. (2007) caution that different model specifications can generate dramatically different estimates. I wanted to ensure any results I obtained from my data surrounding a flipped classroom do not result from an OLS model's unique specification. I therefore used a matching technique to ameliorate concerns for model specification. I "pre-processed" the data following Ho et al. (2007) to match treated observations (students in the flipped classroom) with untreated observations (students in the traditional classroom) based on propensity scores for being in a flipped classroom. This technique generates a sample of observations that are very similar, except for the treatment of interest. It is often the case multiple students in the treatment and non-treatment group have the same propensity scores. I therefore used full matching, which keeps all observations in the non-treatment group with the same propensity score as the treatment group observations (Hansen, 2004).

The tally after pre-processing was 67 propensity-score matches out of the 83 students in the sample. I then used this matched sample to model student outcomes through OLS with robust standard errors.

Results and Discussion

TABLE I
Analysis of Undergraduate Student Performance on Applied Statistics Problems, Department of Political Science, Boise State University, 2012, using Propensity-Score Matching.

	Relevance, Hypotheses as DV	Variables, Methodology as DV	Evidence, Diagnostics as DV	Implication, Conclusions as DV	Clarity, Organization as DV	Overall Score as DV
Independent Variables						
Flipped Classroom	1.06 0.68	1.93 0.75*	4.57 0.30**	3.64 0.19**	-2.29 1.82	1.46 0.01**
Time Since Pre- Requisite	-0.53 0.46	-0.31 0.12*	-0.60 0.22*	-0.58 0.27*	-0.73 0.58	-3.56 0.20**
Pre-Req. in Pol. Sci. Dept.	0.66 0.29*	2.08 0.85*	2.94 0.43**	1.62 0.58*	1.37 0.16**	3.76 0.30**
Gender	4.39 0.46*	2.56 1.44	-1.05 0.49*	1.27 0.83	0.35 0.11*	2.54 0.86*
GPA	5.43 1.96*	6.72 2.01*	6.94 1.47**	7.80 1.61**	2.48 0.52**	6.35 0.59**
Constant	15.95 4.68*	14.82 6.81*	12.43 2.89**	13.49 2.17**	14.70 4.23*	72.53 8.77**
N	67	67	67	67	67	67
Adjusted R ²	0.65	0.61	0.80	0.73	0.59	0.68

Standard errors in parentheses are robust. *p < .05; **p < .01

The following results emerge from my regression models: First, students in the flipped classroom performed slightly better on their final projects than students in the traditional classroom. The differences are statistically significant, but the magnitude of the differences is small: I estimate the practical impact of a flipped classroom to be only one and a half points on a 100 point scale. In other words, if the mean project grade in the traditional classroom is an 83, then I estimate students in the flipped classroom will earn a mean score of 84.5 on the same assignment. However, disaggregating the final project data demonstrates the ways a flipped classroom still influenced student performance considerably. Students in the flipped classroom exhibited statistically significant performance advantages over those in the traditional classroom on several elements of their final projects. Flipped classroom students scored higher on the sections surrounding variable identification and description of methodology, presentation of evidence from regression models and regression diagnostics and the implications of the results/conclusions. The coefficients on the flipped classroom variable indicate estimated impacts of approximately 2, 4.5 and 3.5 points on each section's 20-point scale, while holding all other variables held constant at their means. These sections of the project are consistently the most difficult for students and mean scores are low: 15.14, 12.79 and 13.92 out of 20, respectively. Yet I had the time to answer all of students' questions directly and conveniently in the flipped classroom, but could spare much less time to answering research-paper questions in the traditional classroom. Students in the flipped classroom thus spent relatively more time receiving instruction and applying their knowledge on these topics compared to students in the traditional classroom. Student performance improved as a

result, in keeping with previous literature on the topic (Lage et al., 2000; Hake, 1998; Crouch and Mazur, 2001; Moravec et al., 2010; DesLauriers et al., 2011). In contrast, I devoted no additional time to discussions of the relevance of students' projects or advice for organizing a research paper in the flipped classroom compared to the traditional one. Subsequently, it is not surprising students in the flipped classroom performed no better or worse than students in the traditional classroom in these areas.

There are several other statistically significant results worth mentioning. First, students who completed the prerequisite course relatively recently and in the Political Science Department performed better than those that took the course long ago and/or outside of the department. On average, students in my sample lost five points (on a 100 point scale) for every additional semester they waited between the introductory and advanced classes. Second, female students outperformed male students at statistically significant rates in many areas. Finally, students with higher GPAs performed better than students with lower GPAs.

Next, teaching evaluations for the two sections demonstrate important differences between student perceptions of flipped classroom vs. traditional course delivery: students in the flipped classroom rated the class and instructor higher than students in the traditional classroom in their end-of-term evaluations. Furthermore, students in the flipped classroom *felt* they learned more than students in the traditional classroom. I use difference of means tests (2-tailed t-test) to determine whether differences between mean scores are statistically significant and unlikely to result from random fluctuations in the data. I present these comparisons below in Table II.

Table II
Teaching Evaluations from Flipped vs. Traditional Classrooms
(mean scores are out of a possible five points, scores closer to 5 reflect positive evaluations)

Component Evaluated	Traditional Classroom	Flipped Classroom
Total Course Evaluation (mean out of 5)	4.38	4.56*
Instructor Mean	4.52	4.60*
Course Mean	4.27	4.33
Self-Assessment of Learning in the Course	4.30	4.51**
Interest in taking Additional Political Science Methodology Course	26%	58%**
Number of Students	40	43

*p < .05; **p < .01

Mean teaching evaluations were 4.38 out of 5 in the traditional classroom compared to 4.56 out of 5 in the flipped classroom. These differences are statistically significant at the .01 level and similar to the mean difference between my evaluations for upper division courses, which are usually electives in my research area, feature relatively fewer students and survey course designs (4.7 mean scores) and my lower-division, large lecture courses (4.45 mean scores). These scores reflect strong teaching evaluations, but I believe I do a much better job imparting the material to students in the upper-division classes than in the lower division courses. Of course, students select into the upper division courses because they enjoy the subject matter or they like my teaching style. I am therefore not surprised by such differences in evaluations between the lower and upper division classes. However, similar differences between one section of a required course and the next are very surprising when I teach both sections. Furthermore, these results are in contrast to the experience described in the Harvard Physics Department in Berrett (2012) where “the average score of a student evaluation in a flipped course is about half of what the same professor gets when using a traditional lecture”. It is therefore all the more striking that my students rated the flipped classroom as superior to the traditional classroom.

Finally, student responses also suggest the flipped classroom's influence on students' desire to learn more in the subject area. The results in Table II show students in the flipped classroom are more interested in taking additional Political Science methodology classes than those in the traditional classroom. 58% of students in the flipped classroom expressed an interest in taking additional methodology classes if they were offered whereas only 26% of those in the traditional classroom said they would enroll in these classes.

Results in favor of affective responses to the flipped classroom technique are important not just because I want students to enjoy the class, but for *why* I want them to enjoy the class. Increasing student enjoyment is a worthy goal because students' sense of how useful, feasible and enjoyable a set of skills are to deploy is likely to translate into future willingness/openness to the possibility of deploying these skills in the future or supplementing these skills so as to use them in an applied fashion (Krathwohl et al., 1973). In this case, students in the flipped classroom may have a greater relative willingness to pursue additional training in statistics and greater confidence using quantitative data analysis in the workplace compared to those in my traditional classroom. It is important to note I am not advocating high grades for all in the hope students will remember the class fondly and revisit the material in the future. Instead, my point is instructional delivery may support learning goals even if it does not directly improve classroom performance.

Unfortunately, the questions on my teaching evaluations are not specific enough for me to test direct hypotheses as to what drove students' higher teaching evaluations in the flipped versus traditional classroom. It is possible students found the flipped classroom environment more comfortable or enjoyable than the traditional classroom for a wide variety of reasons, including more efficient use of time, perceived improvement in learning outcomes or camaraderie with classmates. Greater specificity in teaching evaluation questions would allow me to evaluate these possibilities in future research and use the results to better understand precisely what drives affective responses in the classroom.

Robustness Checks

I took several additional steps to ensure the reliability and validity of the quasi-experiment. First the statistical estimates I obtained in Table I are similar using different estimation techniques. I supplemented propensity-score matching with a model using ordinary least squares and robust standard errors as well as a model using OLS with robust standard errors on bootstrap simulated data. The results of estimation using these models appear in Tables Ia and IIa of the Technical Appendix. I also present data on the correlations among variables and variance inflation factors to allay concerns for multicollinearity in Table IIIa. Finally, I had never taught the lower division prerequisite to the advanced statistics class prior to flipping the classroom. The results on the prerequisite variable (time since prerequisite, place prerequisite was taken) therefore cannot reflect a performance bump for students who have taken both classes in sequence with me as an instructor.

Limitations of the Research

The chief limitation of this approach is the reliance on dependent variables (grades) I assigned to my own students. I can claim to be impartial in my assignment of these grades across treatment and control section, but there are always opportunities for subconsciously-skewed grading and biased results when one administers the treatment, knows which subjects are in the treatment group and evaluates the treatment results. For instance, the known publication bias in favor of positive results could lead to differences in grading or in how I approached students across sections. I could have subconsciously inflated grades, devoted more office-hour time to students in the treatment section or provided the treatment section with any number of other subtle, subconscious advantages. However, I do not think data from the treatment section suffers from grade inflation or preferential instructor behavior for several reasons. First, my university and department keep a careful watch on students' scores and place pressure on instructors to maintain discerning grade distributions. I am sensitive to these pressures as an untenured faculty member and take great care to avoid the appearance of easy grading in exchange for higher teaching evaluations. I therefore made a regular, conscious effort to take stock of how I was behaving across the two sections and how the grades were distributed throughout the study. Second, one would expect any pro-treatment bias to result in higher grades across the course as well as across individual assignments. Yet, this is not what I find. Instead, scores in the treatment section are only higher in certain specific areas, not overall project grades. I thus have some evidence against a personal bias generating the results, though the design would still be stronger if I had not known who was in the treatment and control groups and had not evaluated the final projects myself.

The relatively small sample of students whose data I captured during a relatively short timeframe represents another limitation of my research. It would be a mistake to assess the general utility of flipping the classroom or any other instructional technique from such a small study alone. Similar data collection in different institutions, with different class sizes and student backgrounds over a longer period of time is necessary to evaluate whether a flipped classroom improves general student performance.

Transferability and Conclusions

Despite the limitations described above, many of my results are *potentially* quite relevant for students at other universities- even to students in other disciplines. First, I suspect students who spend more time working with material in a guided classroom environment will outperform students who spend less time with the material, all things being equal. A flipped classroom ensures students actively engage the material during class and also remain for question and answer sessions. Normally, only the most motivated students attend office hours to ask the questions that arise regularly in a flipped classroom. Consequently, students who are less motivated still benefit from other students' questions in a flipped classroom when I correct everyone's misconceptions at the time of need, rather than potentially weeks later after I receive enough evidence from the results of an exam. The result is higher performance for students in a flipped classroom on *areas emphasized in that classroom*. A flipped classroom is not a panacea curing all educational ills. Class time is still limited, instruction is never perfect and students can still perform poorly on assignments for a wide variety of reasons, including lack of instruction on a specific element of the assignment. Nevertheless, my research adds to the body of scholarship highlighting the ways flipping the classroom can provide additional, productive class time and improve student performance.

Second, students' greater enjoyment and greater self-assessed learning outcomes in the flipped classroom are potentially generalizable to a broader population. Students enjoy "homework" more when it is done in class instead (Cox and Singer, 2011, Bempechat et al., 2011). Students everywhere exhibit affective responses to material- they learn more when they enjoy the material, are likely to revisit the subject in the future and use it in practice because they have a positive association with it. The higher levels of enjoyment and self-assessed learning for students in the flipped classroom may thus make it more likely these students will use their newfound skills compared to students in the traditional class, all without undermining learning outcomes.

There is a deeper pedagogical question associated with a flipped classroom that extends past student performance on semester-long projects: does flipping the classroom privilege short-term outcomes over long-term goals for educational development? I see potential for a flipped classroom to improve students' command of the material at hand, but also to create over-reliance on instructors to lead students through the material and impede learning in the long term. However, a flipped classroom can potentially help students achieve both short-term and long-term learning goals so long as instructors take care to treat a classroom flip as a form of scaffolding. Devoting more classroom time to activities requiring higher-order cognitive thinking will reap rewards at the beginning of a course, but the technique can still promote students' ability to teach each other and themselves if class activities are performed with less and less direct instructor support as the semester progresses. Ultimately, my research has helped structure my course delivery in this area and improve student outcomes. The kind of data I collected to model these outcomes is usually available, which should make it possible for others to do the same.

Acknowledgements

I would like to thank Boise State University's Center for Teaching and Learning for providing extensive support for this research. In particular, I would like to thank the Director of the Center, Dr. Susan Shadle, along with other participants in the Boise State Early Career Teaching Scholars Program. Drs. Yong Gao, Matt Genuchi, Jill Heney and Casey Keck provided valuable insight and suggestions throughout the research process. Finally, I am grateful to the Political Science Department at Boise State University for affording me the opportunity to experiment in the classroom and supporting scholarship on teaching and learning.

Biographical Sketch

Michael Touchton is an Assistant Professor of Political Science at Boise State University. He studies the comparative political economy of development and underdevelopment- in particular the relationship between states and markets around the world. His current book project explores the ways democracies make credible commitments to property rights, contract enforcement and the rule of law.

References

- Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching*. John Wiley & Sons.
- Anderson, L., Krathwohl, D., & Bloom, B. (2005). *A taxonomy for learning, teaching, and assessing*. Longman.
- Baldwin, T., Bedell, M., & Johnson, J. (1997). The social fabric of a team-based MBA program: Network effects on student satisfaction and performance. *Academy of Management Journal*, 40(6), 1369-1397.
- Battistich, V., Solomon, D., Kim, D. I., Watson, M., & Schaps, E. (1995). Schools as communities, poverty levels of student populations, and students' attitudes, motives, and performance: A multilevel analysis. *American Educational Research Journal*, 32(3), 627-658.
- Bempechat, J., Li, J., Neier, S. M., Gillis, C. A., & Holloway, S. D. (2011). The homework experience: Perceptions of low-income youth. *Journal of Advanced Academics*, 22(2), 250-278.
- Bergmann, J., & Sams, A. (2012). *Flip Your Classroom. Reach Every Student in Every Class Every Day*. International Society for Technology in Education.
- Berrett, D. (2012). How 'Flipping' the Classroom Can Improve the Traditional Lecture. *Chronicle of Higher Education*, 12.
- Conger, D., & Long, M. (2010). Why are men falling behind? Gender gaps in college performance and persistence. *The Annals of the American Academy of Political and Social Science*, 627(1), 184-214.
- Cox, T., & Singer, S. (2011). Taking the Work out of Homework. *Mathematics Teacher*, 104 (7), 514-519.
- Crouch, C., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69, 970.
- DesLauriers L, Schelew E, & Wieman C. (2011). Improved learning in a large-enrollment physics class. *Science*, 332(6031), 862-864.
- Fulton, K. (2012). Upside Down and Inside Out: Flip Your Classroom to Improve Student Learning. *Learning & Leading with Technology*, June-July, 12-17.
- Hake, R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64-74.
- Hansen, B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association*, 99(467), 609-618.
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, 15(3), 199-236.
- King, A. (1993). From sage on the stage to guide on the side. *College teaching*, 41(1), 30-35.
- Krathwohl, D. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice*, 41(4), 212-218.
- Lage, M. J., Platt, G. J., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an inclusive learning environment. *Journal of Economic Education*, 31(1), 30-43.
- Mazur, E. (2009). Farewell, Lecture? *Science* 323: 50-51.
- Moravec, M., Williams, A., Aguilar-Roca, N., & O'Dowd, D. (2010). Learn before Lecture: A Strategy That Improves Learning Outcomes in a Large Introductory Biology Class. *CBE—Life Sciences Education*, 9 (4), 473-81.
- Sahin, A., Ozcan A., & Niyazi E. (2012). How Computer Course Enrollment is Related with Students STEM Major Choice, AP Taking, SAT Scores, Gender and Socio Economic Level. *Society for Information Technology & Teacher Education International Conference*. 2012 (1).

Technical Appendix: Flipping the Classroom and Student Performance in Advanced Statistics: Evidence from a Quasi-Experiment

Table Ia

Analysis of Undergraduate Student Performance on Applied Statistics Problems, Department of Political Science, Boise State University, 2012, using OLS with Robust Standard Errors

Standard errors in parentheses are robust. *p < .05; **p < .01

Independent Variables	Relevance, Hypotheses as DV	Variables, Methodology as DV	Evidence, Diagnostics as DV	Implication, Conclusions as DV	Clarity, Organization as DV	Overall Score as DV
Flipped Classroom	1.30 0.74	2.04 1.32	4.36 2.08*	3.29 1.26*	-1.88 1.14	1.17 0.21**
Time Since Pre-Requisite	-0.52 0.13**	-0.61 0.10**	-1.09 0.23**	-0.76 0.18**	-0.35 0.07**	-4.76 0.41**
Pre-Req. in Pol. Sci. Dept.	0.79 0.16**	2.46 0.38**	2.71 0.40**	1.83 0.53*	1.66 0.34**	3.95 0.32**
Gender	5.22 0.18**	3.46 0.63**	-2.51 0.68**	1.35 0.61*	0.70 0.54	3.08 0.93*
GPA	5.66 2.15*	7.03 2.34*	6.93 2.28*	7.72 1.98**	2.58 0.36**	7.32 0.26**
Constant	16.04 2.46**	14.96 2.08**	12.61 1.90**	13.57 2.25**	14.89 2.53**	64.80 2.61**
N	83	83	83	83	83	83
Adjusted R ²	0.65	0.60	0.66	0.68	0.52	0.63

Harrell (2001) contends one should strive for at least 10-20 observations per covariate in order to make “reasonable-size estimates with reasonable power” using OLS regression. My data meets the threshold of 10, but not of 20 observations per covariate. I supplement the first two modeling techniques with a bootstrap simulation of additional data given the concern for the reliability of estimates generated from similarly small samples. This technique resamples data with replacement to produce a new, larger dataset drawn from the original sample. The hope is this new dataset allows for more accurate inferences and estimates relative to the original models using the smaller dataset (Efron and Tibshirani, 1993). I then use OLS with robust standard errors to re-estimate the relationship between instructional design, student characteristics and performance on applied statistics projects.

Table IIa
 Analysis of Undergraduate Student Performance on Applied Statistics Problems, Department of Political Science,
 Boise State University, 2012, using OLS with robust standard errors on Bootstrap-Simulated Data.

	Relevance, Hypotheses as DV	Variables, Methodology as DV	Evidence, Diagnostics as DV	Implication, Conclusions as DV	Clarity, Organization as DV	Overall Score as DV
Independent Variables						
Flipped Classroom	1.16 0.59	1.89 0.33**	4.34 0.38**	3.85 0.31**	-1.68 0.72*	1.36 0.01**
Time Since Pre- Requisite	-0.46 0.29	-0.31 0.12*	-0.50 0.23*	-0.36 0.15*	-0.67 0.35	-5.72 0.01**
Pre-Req. in Pol. Sci. Dept.	0.51 0.42	2.14 0.39**	2.87 0.55**	1.46 0.64*	1.43 0.58*	4.06 0.09**
Gender	3.84 2.33	2.63 1.75	-0.91 0.57	1.46 0.71*	0.58 0.24*	2.93 1.25*
GPA	5.62 2.18*	6.94 1.39**	6.80 1.46**	8.05 0.84**	2.63 0.66**	8.03 0.04**
Constant	15.25 6.18*	14.69 5.37*	12.85 3.61*	13.26 3.48**	15.01 5.73*	67.15 3.93**
N	12,370	12,370	12,370	12,370	12,370	12,370
Adjusted R ²	0.62	0.55	0.60	0.64	0.57	0.57

Standard errors in parentheses are robust. *p < .05; **p < .01

High levels of multicollinearity are not likely among the independent variables in my models due to the low correlation among them along with the statistically significant relationships they all exhibit with the dependent variable. This information is presented below in Table IIIa. Furthermore, all variance inflation factors are below four, indicating little concern for multicollinearity based on Pan and Jackson's (2007) more-conservative modification of the standard in the discipline of a VIF of 5.

Table IIIa
Correlelogram of Independent Variables in Table I

	Flipped Classroom	Time Since PreReq	PreReq in Pol. Sci.	Gender	GPA
Flipped Classroom	1.00				
Time Since PreReq	0.07	1.00			
PreReq in Pol. Sci.	-0.09	-0.24	1.00		
Gender	0.11	-0.14	0.26	1.00	
GPA	0.18	-0.21	0.19	0.34	1.00

Instructions for Final Research Paper

A final, double-spaced, 15-25 page report is due by email on December 12th, by midnight. Late reports will have five points deducted for each additional day they are late. The report should include at least the following items:

1. The research question investigated and a discussion of its importance (for scholars, for policymakers, etc.).
2. A review of the literature on your subject (this should probably be three or four pages)
3. A description of the dependent variable including its operationalization.
4. The hypotheses you will test and a description of how you generated them (this should be at least two pages).
5. The equation you estimated as well as a justification for your model specification (this can be a paragraph or two depending on how extensive #4 was).
6. Identification of the independent variables and how they were operationalized. For example, how is each variable being measured? Where did this data come from? What are the advantages/disadvantages of these measures?
7. The results of the multiple regression estimation (the tables from SPSS)
8. Discussion of results with reference to the tables. Interpret each element SPSS reports for individual independent variables as well as the model as a whole.
9. Discussion of the regression assumptions (autocorrelation, heteroskedasticity, linearity, multicollinearity, normality, residual outliers). There is a lot to discuss here, but you can do a lot of it based on visual assessments using scatter-plots and histograms of the data.
10. Assessment of the model, including its validity, how well supported the hypotheses were by the results, and the implications of the results for future research. What do your results mean for policy makers? For politicians? For politically active groups? etc. This section should be at least three or four pages.
11. Bibliography (this does not count toward page total). There is no set number of sources, but a good literature review will have a minimum of three sources per variable.

Grading Rubric for Final Projects

The grade you earn on your paper will depend on the extent to which your paper answers the questions included within each category below. Each section is worth 20% of the total.

1. Explanation of the project's central question, its significance and hypotheses (20%)

(20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0)

What is your central issue? Why is it important? What does the literature say about your topic? What are your hypotheses surrounding the relationships between these variables? What evidence or arguments led you to formulate these null and alternate hypotheses?

2. Description of variables and methodology employed- 20%

(20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0)

Describe your independent and dependent variables. Why did you select these variables? How are they measured? Are there gaps between concepts and measurements? What can multiple regression analysis using this data tell you about relationships among your variables? What does regression analysis leave out?

3. Presentation of evidence/diagnostics- 20%

(20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0)

Interpret the results of regression analysis and describe the relationships (if any) you found in your data. Do you have reason to believe your data violates the assumptions surrounding regression analysis? Present evidence and discuss the extent to which your data violates or does not violate OLS regression assumptions.

4. Implications/Conclusion- 20%

(20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0)

What did your research uncover? What does this lead you to believe with regard to previous scholarship on the topic? What are the implications for policy in your area? Do you need more information to address the issue? Why/What type?

5. Overall Clarity/Organization- 20%

(20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0)

Does the paper clearly describe the various aspects of the project? Did you present these aspects at the appropriate time?