The better blend? Flipping the principles of microeconomics classroom

Neal H. Olitsky, Sarah B. Cosgrove*
University of Massachusetts Dartmouth, Department of Economics, 285 Old Westport Rd, N Dartmouth, MA 02747, USA

A R T I C L E   I N F O
Article history:
Received 25 June 2015
Received in revised form 30 September 2015
Accepted 9 October 2015
Available online 5 November 2015

JEL codes:
A22

Keywords:
Blended learning
Flipped classroom
Propensity score
Matching estimator

A B S T R A C T
This study provides a comprehensive analysis of an experiment that attempts to cut costs and improve learning outcomes through a “flipped-blended” class. We discern effects of this pedagogy on learning outcomes in principles of microeconomics courses. We control for key background variables and use differences-in-differences with a matching estimator to test whether students in the flipped-blended classes learn economic concepts and tools better than students in classes with interactive lectures and online homework, but no online lectures. Findings suggest that average student improvement is significantly higher in flipped classes than in non-flipped classes, though the difference in improvement is modest.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Amidst increasing pressure to achieve higher levels of measurable student learning with fewer resources, faculty in higher education are redesigning their curricula and/or course delivery to incorporate technology. Among the strategies gaining popularity are blending and flipping. Blending is commonly defined as redesigning a class such that part of the coursework is completed online and the remainder occurs in a traditional face-to-face setting. Because less classroom time is required, blending has the potential to reduce operating costs for universities. Flipping is inverting the activities traditionally conducted during face-to-face class time with activities students complete after class.

Blending is becoming a common delivery mode in higher education because of its success in achieving student learning outcomes while increasing flexibility for students. While data on the prevalence of blending are scant, “it is generally believed that blended learning has reached well into the mainstream of American higher education” (Picciano et al., 2013, 3)1 and one survey found that 80% of all higher education institutions and 93% of doctoral institutions offer hybrid or blended learning courses (Arabasz et al., 2003). The success of blended classes is largely attributable to increased student engagement with the material and expanded access to multi-media offerings. However, students in blended classes in economics have not shown significant improvements in achievement compared to students in traditional classes (Brown and Liedholm, 2002; Terry and Lever, 2003; Olitsky and Cosgrove, 2013; Cosgrove and Olitsky, 2014). Moreover, students in

* Corresponding author.
E-mail addresses: neal.olitsky@umassd.edu (N.H. Olitsky), sarah.cosgrove@umassd.edu (S.B. Cosgrove).

1 Picciano et al. (2013) offers several explanations for the lack of data on blending including; many faculty are blending but do not label the mode as such, colleges and universities collect data on fully online courses but not blended courses, and there is no single widely accepted definition of blending.

http://dx.doi.org/10.1016/j.iree.2015.10.004
1477-3880/ © 2015 Elsevier Ltd. All rights reserved.
blended classes have shown lower levels of knowledge retention than students in traditional classes (Cosgrove and Olitsky, 2014).

Another pedagogical approach that is becoming more prevalent is “flipping” the classroom. With this approach, students are introduced to course material before attending a face-to-face class on that material (Walvoord and Anderson, 2011), using class time to achieve higher levels of learning. Flipping goes by other names, such as “inverting the classroom” (Lage et al., 2000) and “peer instruction” (Crouch and Mazur, 2001) and can take different forms, but it is an approach under which introduction to the material that is traditionally done in the classroom is completed by the students independently before the class and the traditional homework is done during class time. Under this approach, students become familiar with the material and achieve the low levels of Bloom’s taxonomy (Anderson et al., 1994), knowledge and some comprehension, before attending class, leaving class time to focus on the more difficult tasks of applying, analyzing, synthesizing, and evaluating. This approach has been shown to be highly effective at improving student achievement, particularly in quantitative subjects such as math and physics (Hake, 1998; Crouch and Mazur, 2001; Deslauriers et al., 2011).

We conduct an experiment that attempts to simultaneously cut costs and improve student achievement in principles of microeconomics classes by combining blending and flipping. The goal is to achieve the gains from blending found in other disciplines that have eluded economics thus far by flipping the course and blending the course more aggressively while also reducing the costs of course delivery. The cost reduction is achieved by doubling the enrollment in the flipped class without changing the teaching load or requiring additional classroom resources.

This study advances the literature in three ways. First, it is the only experimental study comparing a flipped-blended class to a modestly blended class. Previous studies showed no significant gains (or losses) in student learning outcomes from modestly blended principles of economics classes. We hypothesize that the degree to which a class is blended may affect the outcomes. The combination of flipping and blending allows us to design the class with the more basic work taking place online and the most challenging work conducted in a face-to-face setting. Second, our methodology controls for selection bias, found in previous studies of online versus face-to-face learning, using differences-in-differences with a matching estimator developed in Heckman et al. (1998). This method allows us to estimate the relationship between flipped coursework and changes in student achievement, controlling both for individual characteristics and for differences between the flipped and non-flipped classes. Third, our data include numerous control variables lacking in the means-only analysis presented from the math and physics disciplines, providing, to our knowledge, one of the only controlled studies on flipped coursework.

The results show no evidence that students in flipped-blended classes perform worse than their non-flipped modestly blended counterparts, despite having less face-to-face class time during the semester, and suggest that students in flipped classes do significantly better. Overall, students in the flipped classes showed a larger improvement between the pretest and the posttest, improving their score on average by 7% points more than students in the non-flipped class. Because students in flipped classes show larger differential gains than those in the control group and because flipped-blended classes can accommodate twice as many students as a traditional course, our results support the implementation of the flipped-blended format as a cost-reducing, outcome-improving alternative to the traditional face-to-face course.

The paper proceeds as follows. Section 2 provides some background, placing the present study in the context of the existing literature concerning blended and flipped course work. Section 3 describes the design of our experiment. Sections 4 and 5 present, respectively, the empirical strategy and a descriptive analysis of the collected data. Section 6 presents the estimation results, and Section 7 discusses the results in the context of the extant literature and Section 8 concludes.

2. Background and literature review

In recent years, there has been a shift toward more blended and fully online learning in higher education. Traditional lecture-based courses are being replaced with varying degrees of online learning, requiring less physical capacity from the universities and providing more flexible schedules to students (Allen et al., 2007). Faculty and students alike are finding ways to leverage technology to improve their effectiveness and efficiency (Goldstein and Katz, 2005; Roblyer et al., 2010). Blending has been shown to be effective at achieving student learning outcomes and increasing course completion rates. Vaughan (2007) provides a thorough survey of the literature on blending, citing a large-scale program for course redesign to a blended format. Of the 30 participating institutions, 20 reported improved learning outcomes and 10 reported no significant difference. Moreover, 75% of the participating institutions that measured changes in drops, failures, and withdrawals saw a reduction compared to the face-to-face classes.

This shift comes at a time when public funding for higher education is declining and enrollment is increasing. While public colleges, universities, and community colleges enroll over 70% of all college students, state funding for these institutions decreased between AY1990-91 and AY2009-10 by “an average of 26% in real terms—even as operating costs increased” (Mettler 2014). Moreover, from 1990 to 2012, enrollment in degree granting post-secondary institutions increased by more than 49% (National Center for Education Statistics, 2013). Colleges and universities are attempting to teach more students in larger classes to reduce the need for classroom space and faculty hours. Further fueling the shift toward blending, some large meta-analyses concluded that blending leads to better student learning outcomes. A major study conducted by the US Department of Education found evidence of better achievement of student learning outcomes in blended and online environments, with a larger advantage in blended classes relative to face-to-face than in fully online classes (Means et al., 2009).
Similarly, research on flipping the class shows remarkable improvements in student learning. For example, Deslauriers et al. (2011) flipped one of two physics classes during the 12th week of the semester. During the first class following the flipped week, students in the flipped and control classes were given a test on the content from that week. The mean scores were 41% in the control class and 74% in the flipped class for an improvement of 2.5 standard deviations. However, because this study did not include any control variables, it is possible that these gains were a result of differences in student characteristics across the treatment and control groups. Another study, Crouch and Mazur (2001), used a modified version of flipping they call “peer instruction” in which questions were posed to the class and responses were recorded using personal response devices (“clickers”). Students in a physics class were then asked to find a classmate who selected a different answer and convince them that their answer was correct. In addition to this peer instruction approach, quizzes were used at the beginning of the class to motivate students to prepare. Comparing results on standardized physics exams like the Mechanics Baseline Test, the authors reported a mean increase from 66% in 1990, the last year in which peer instruction was not used, to 72% in 1991 and 79% in 1997 after peer instruction. Once again, no control variables were included in the analysis. Finally, Hake (1998) saw significant learning gains of almost two standard deviations when he flipped his physics class. While these studies lack the empirical rigor necessary for strong conclusions to be drawn about the technique, their results suggest that the flipped approach is certainly worth considering. Berrett (2012) succinctly stated the potential value of flipping as: “The immediacy of teaching in this way enables students’ misconceptions to be corrected before they emerge on a midterm or final exam”.

Despite the strong evidence reported in other disciplines, very few studies describe examples of flipping or partial flipping in economics classes. Lage et al. (2000) recount flipping in an attempt to address multiple learning styles. While they report positive student and faculty perceptions of the course, they do not empirically test for differences in student learning outcomes. Roach (2014) also focuses on student perceptions of flipping but reports secondarily that students in a partially flipped class performed better on a common final exam, compared to students in traditional classes. Roach cautions that no control variables are included in the analysis. Calimeris and Sauer (2014) compare test scores of students in a traditional principles of microeconomics course to those of students in a flipped class. Their results from OLS regressions show that students in the flipped class performed significantly better than their counterparts in the traditional class on the midterm and final exams.

While they do not engage in flipping, three studies report relevant research in principles of microeconomics classes. First, Figlio et al. (2013) compare test scores for students who participated in live lectures with those of students who had access to recorded lectures posted on the internet, but the classes were otherwise identical. They find that average test scores for students in the live class were modestly higher than for students in online classes. Furthermore, Hispanic students, male students, and students with lower than average grade point averages (GPAs) particularly benefitted from live lectures. Second, Chen and Lin (2012) study the effect of making online recorded lectures available to students. They find that the use of these supplemental video lectures results in an average improvement of four percentage points on exams. Finally, Emerson and Taylor (2004) study the effects of using experiments during some class time and compare student performance to control classes which rely more heavily on traditional lectures. They find that students in classes with interactive experiments improved their scores on the Test of Understanding in College Economics (TUCE) by an average of 2.42–2.99 points over the control group. Moreover, they find disproportional gains to low achievement students and females from the

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental design of control and treatment groups.</td>
</tr>
<tr>
<td>Control group</td>
</tr>
<tr>
<td># of Face-to-face (F2F) classes</td>
</tr>
<tr>
<td># of F2F hours of instruction</td>
</tr>
<tr>
<td># of F2F classes offered per week</td>
</tr>
<tr>
<td>Timing of data collection</td>
</tr>
<tr>
<td># of Students present for F2F classes</td>
</tr>
<tr>
<td>Pre F2F expectations</td>
</tr>
<tr>
<td>F2F class activities</td>
</tr>
<tr>
<td>Post F2F expectations</td>
</tr>
<tr>
<td>Undergraduate teaching aides</td>
</tr>
<tr>
<td>Maximum of 60</td>
</tr>
<tr>
<td>Watch and take notes on video lectures</td>
</tr>
<tr>
<td>Complete brief low-stakes multiple choice quiz on paper</td>
</tr>
<tr>
<td>Small group activities (90% of time)</td>
</tr>
<tr>
<td>Complete follow-up assignment(s) on online homework system</td>
</tr>
<tr>
<td>The same two upperclassmen assisted the treatment group throughout the treatment period</td>
</tr>
</tbody>
</table>
experimental curriculum. While it seems that online lectures are not a good substitute for face-to-face lectures, they do seem to be a good complement. In addition, substituting some lecture time with engaged learning activities aids learning. Collectively, this evidence appears favorable for flipping in economics.

3. Experiment design

This experiment compares assessment results from students in modestly blended classes with results from students in an aggressively blended, flipped class. Our sample consists of assessment results from three principles of microeconomics classes across two semesters. The control group is comprised of two 60-student classes, one in the fall 2013 semester and one in the spring 2014 semester, scheduled to meet on Tuesdays and Thursdays for 75 min each day. The treatment group is one 120-student flipped-blended class from the spring 2014 semester that was also scheduled to meet on Tuesdays and Thursdays for 75 min each day. All three of the observed classes were listed as “blended” in the schedule of classes so students were unaware of any difference in the classes at the time of registration. While students could have switched from one of the spring 2014 classes to the other after learning of the format during the first week of classes, none of the students who dropped the non-flipped class enrolled in the flipped class and none of the students who dropped the flipped class enrolled in the non-flipped class. It is also important to note that principles of microeconomics is a required class for both economics and business majors, but also covers a general education requirement for all students at the university. Unlike other campuses, there is no prescribed order in which students are required or encouraged to enroll in principles of microeconomics and principles of macroeconomics. Thus, students enrolling in the spring classes should not be considered to be “off track” as compared to students enrolling in the fall class, or vice versa. Only one student enrolled in the course during the spring 2013 semester and repeated the course during the fall 2014 semester. This observation was excluded from the study. An outline of the differences between the control and treatment groups is available in Table 1 and described below.

Students in the control classes met face-to-face for 19 classes (23.75 h) or 83% of the scheduled classes and had online activities in lieu of class for the other 17% of the scheduled classes, excluding test days. These students were expected to read the relevant textbook chapter(s) prior to attending their face-to-face class for that week. The face-to-face meetings consisted primarily of interactive lectures, with limited pair-and-share exercises and small group work totaling approximately 10% of face-to-face class time. The online activities were discussion boards which involved researching and writing about examples of topics discussed in class (e.g., tradeoffs between equity and efficiency, externalities) and creating and posting original problems akin to those covered in class (e.g., specialization and trade, production and cost calculations). There were no undergraduate teaching aides for the control classes.

The treatment class met face-to-face for 14 classes (17.5 h) or approximately 61% of the scheduled classes and had online activities in lieu of class for the other 39% of the scheduled classes, excluding test days. This class was divided randomly into two groups with 60 students in each group. One group attended face-to-face classes on Tuesdays and worked online the remainder of each week, while the other group attended face-to-face classes on Thursdays and worked online the remainder of each week. These students were expected to read the relevant textbook chapter(s), and view, study, and take notes on pre-recorded video lectures prior to attending their face-to-face class for that week. They completed a low stakes 6–10 question multiple-choice quiz at the beginning of the class to incentivize preparation for class. Mid-semester, students were permitted to use their notes when taking the quizzes to provide additional incentive to prepare for class. The videos were made by an Associate Professor with more than ten years of experience teaching this course using Camtasia Relay software to capture audio and corresponding writing on a tablet PC using OneNote.

During class time, students in the treatment class worked in randomly assigned groups of three or four on problems of increasing difficulty covering the topics in that week’s videos. Students were randomly allocated into different groups during each class period so that students with different abilities and motivation levels were mixed. Some of the questions were in multiple-choice format, consistent with the quizzes and unit tests, and some were open response problems. The majority of the questions required application and analysis of the topics for the week. Additionally, some questions involving evaluation and synthesis were included at the end of the lesson. The professor and two undergraduate teaching aides circulated through the groups answering questions and assisting the students. The same two teaching aides worked with the class for the entire semester. At several points during each class, the whole group was brought together to summarize answers to the problems and to address frequently asked questions.

To best capture the true effect of the curriculum redesign, many aspects of the course were consistent across the control and treatment groups. Both groups were required to use the same textbook and corresponding online homework system with identical assignments and due dates. All lecture material was the same, though the mode of delivery differed between the control and treatment groups. The same instructor taught all of the students in the control and treatment groups. All tests

---

2 This control group matches the format in which no gains from blending were found previously (Olitsky and Cosgrove, 2013; Cosgrove and Olitsky, 2014).
3 To ensure that all students in the flipped class received the same amount of face-to-face time given the schedule of holidays for the semester, excluding test days all 120 students met on the first two days of class and on two classes that immediately preceded unit tests.
4 Certainly, the development of the videos required a substantial time cost; however, we disregard this cost in making our cost-savings argument, as the videos are reusable for future classes at zero marginal cost.
5 Samples of the in-class work are available from the authors upon request.
were administered on the same schedule and were identical across the classes; however, multiple versions of tests were distributed to dissuade cheating.

The assessment data used in this study were collected at two distinct times during the class. All students in both the control and treatment groups were given an identical 25-question pretest on their first day of class. While no credit was awarded for the pretest, students were encouraged to achieve the best score they could. No results of the pretest were distributed to the students, nor were the questions returned to them. All students in both the control and treatment groups were also given the same 25-question test on the last day of class, as a post-test for which extra credit was awarded. Our IRB approval allowed us to notify students on the final day of the class that they were part of an experiment and seek their consent at that time, so as to not influence their behavior during the semester.

The pre/post-test was not general in nature, but rather specific to the course content with questions like those found on a course exam. The pre/post-test included nine questions from the first unit and eight questions from the second and third units, with questions of varying degrees of difficulty. The pre/post-test is not a general test of economic knowledge, but rather a test designed to assess the specific curriculum of the course. The first unit covered the introductory chapters, focusing on the production possibilities curve (chapter 2) comparative advantage and trade (chapter 3), the supply/demand model (chapter 4), and elasticity (chapter 5). The second unit covered public policy and its effect on market outcomes, including government intervention (chapter 6), market efficiency (chapter 7), taxation (chapter 8), externalities (chapter 10), public goods (chapter 11), and introduced the theory of the firm, covering production costs (chapter 13). The third unit focused on market models, covering competitive firms (chapter 14), monopoly (chapter 15), monopolistic competition (chapter 16) and oligopoly (chapter 17).

The collection of assessment data at two points allows us to measure differential changes in learning that occur during the semester. In addition, by separating the analysis into units, we can analyze the possibility that flipping will affect students differently either based on the nature of the material or based on the amount of time they have to adjust to the flipped mode.

4. Methodology

To estimate the effects of flipped coursework, we begin by obtaining the raw differences-in-differences (DID) in performance between students in the flipped class and students in the non-flipped classes. Following conventional notation, let $Y_i$ be the outcome of student $i$. Further, let $D_i$ be the treatment indicator, taking a value of one if individual $i$ had taken a flipped class, and zero otherwise. We analyze the difference in the pre-test and the post-test scores using an indicator variable that takes a value of one if the observation is a post-test score and a value of zero if it is pre-test score; this is reported by the POST variable. Given the aforementioned notation, we obtain the raw DID by estimating

$$Y_i = \beta_0 + \beta_1D_i + \beta_2\text{POST}_i + \beta_3(\text{POST} \times D)_i + \epsilon_i$$

(1)

The next set of models we estimate are augmented DID regressions to account for time-invariant individual characteristics, including demographics, individual characteristics, and academic achievement. The augmented DID model is as follows:

$$Y_i = \beta_0 + \beta_1D_i + \beta_2\text{POST}_i + \beta_3(\text{POST} \times D)_i + \mathbf{X}_i\beta_x + \epsilon_i$$

(2)

where $\mathbf{X}_i$ is a matrix consisting of the aforementioned controls.

One of the key assumptions of the DID estimation procedure is the “common trend” assumption (Angrist and Pischke, 2009). The common trend assumption is the assertion that in the absence of treatment, the change in outcomes would be the same, on average, both for the treatment group and for the control group. In the context of the present study, the common trend assumption states that students in the flipped class would have shown the same change in assessment scores as the non-flipped class, if those in the flipped class were not exposed to the treatment (e.g., the flipping). The common trend assumption can be written as follows (Smith and Todd, 2005).

$$E(Y_{0t} - Y_{0f} | D = 1) = E(Y_{0t} - Y_{0f} | D = 0)$$

(3)

In the above equation $Y_{0t}$ represents the non-treated outcome in the post-treatment period and $Y_{0f}$ represents the non-treated outcome in the pre-treatment period. In addition, because we only observe the flipped outcomes of students in the flipped class, the left hand side of Eq. (3) is a counterfactual that cannot be observed. Further, Angrist and Pischke (2009) note one way to refute or verify the common trend assumption is to examine several pre-treatment periods and examine whether the pretreatment trends look similar between the treatment and control groups. Unfortunately for the present study, there is only one pretreatment observation, making this type of verification impossible. As a result, we rely on an alternate estimation procedure, which weakens the requirement that this assumption holds (Blundell and Costa Dias, 2009).

In addition to the augmented DID model, we employ differences-in-differences with a matching estimator (MDID) presented in Heckman et al. (1998) and discussed in several other studies (Smith and Todd 2005; Caliendo and Kopeinig, 2008; Blundell and Costa Dias 2009). The MDID estimator determines the effect of a treatment by conditioning outcomes on

---

8 For details on our method of computing difficulty and discrimination indices for the questions, please see Appendix A.
Table 2
Descriptive statistics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
<td>Diff</td>
</tr>
<tr>
<td>Pre-test: all (% correct)</td>
<td>28.060</td>
<td>8.970</td>
<td>27.686</td>
<td>9.481</td>
<td>28.449</td>
<td>8.459</td>
<td>−0.763</td>
</tr>
<tr>
<td>Pre-test: Unit 1 (% correct)</td>
<td>32.390</td>
<td>15.630</td>
<td>32.571</td>
<td>16.041</td>
<td>32.200</td>
<td>15.261</td>
<td>0.371</td>
</tr>
<tr>
<td>Post-test: Unit 1 (% correct)</td>
<td>36.560</td>
<td>15.590</td>
<td>23.529</td>
<td>15.161</td>
<td>29.719</td>
<td>15.494</td>
<td>−6.19*</td>
</tr>
<tr>
<td>Post-test: Unit 2 (% correct)</td>
<td>53.150</td>
<td>13.530</td>
<td>56.063</td>
<td>12.961</td>
<td>50.737</td>
<td>13.595</td>
<td>5.327*</td>
</tr>
<tr>
<td>Post-test: Unit 3 (% correct)</td>
<td>48.840</td>
<td>16.950</td>
<td>50.970</td>
<td>15.667</td>
<td>47.076</td>
<td>17.858</td>
<td>3.894</td>
</tr>
<tr>
<td>Post-test: all (% correct)</td>
<td>63.580</td>
<td>20.600</td>
<td>67.659</td>
<td>18.464</td>
<td>60.197</td>
<td>21.767</td>
<td>7.461*</td>
</tr>
<tr>
<td>Flipped</td>
<td>4.87</td>
<td>0.51</td>
<td>5.098</td>
<td>18.715</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td># ECO classes taken</td>
<td>0.475</td>
<td>0.621</td>
<td>0.594</td>
<td>0.662</td>
<td>0.362</td>
<td>0.560</td>
<td>0.232***</td>
</tr>
<tr>
<td>Cumulative GPA</td>
<td>2.818</td>
<td>0.763</td>
<td>2.696</td>
<td>0.843</td>
<td>2.933</td>
<td>0.661</td>
<td>−0.237***</td>
</tr>
<tr>
<td>Cumulative credits</td>
<td>42.426</td>
<td>21.537</td>
<td>39.179</td>
<td>19.435</td>
<td>45.506</td>
<td>22.991</td>
<td>−6.327***</td>
</tr>
<tr>
<td>Respondent is white</td>
<td>0.678</td>
<td>0.468</td>
<td>0.703</td>
<td>0.458</td>
<td>0.655</td>
<td>0.477</td>
<td>0.048</td>
</tr>
<tr>
<td>Respondent is black</td>
<td>0.121</td>
<td>0.327</td>
<td>0.133</td>
<td>0.341</td>
<td>0.109</td>
<td>0.313</td>
<td>0.024</td>
</tr>
<tr>
<td>Respondent is male</td>
<td>0.602</td>
<td>0.490</td>
<td>0.618</td>
<td>0.487</td>
<td>0.586</td>
<td>0.494</td>
<td>0.032</td>
</tr>
<tr>
<td>On-campus resident</td>
<td>0.773</td>
<td>0.420</td>
<td>0.818</td>
<td>0.387</td>
<td>0.730</td>
<td>0.445</td>
<td>0.088</td>
</tr>
<tr>
<td>Composite SAT score</td>
<td>1026.932</td>
<td>124.810</td>
<td>1041.939</td>
<td>126.688</td>
<td>1012.701</td>
<td>121.661</td>
<td>29.238*</td>
</tr>
<tr>
<td>Business major</td>
<td>0.681</td>
<td>0.467</td>
<td>0.721</td>
<td>0.450</td>
<td>0.644</td>
<td>0.480</td>
<td>0.078</td>
</tr>
<tr>
<td>% of F2F classes attended</td>
<td>91.846</td>
<td>12.646</td>
<td>92.733</td>
<td>11.216</td>
<td>91.004</td>
<td>13.847</td>
<td>1.729</td>
</tr>
<tr>
<td>Observations: total</td>
<td>339</td>
<td>165</td>
<td>174</td>
<td>76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations: pre-test</td>
<td>200</td>
<td>102</td>
<td>98</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations: post-test</td>
<td>139</td>
<td>63</td>
<td>76</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 7 | For a detailed explanation of the MDID estimation, please see Appendix A. | (1) (3) – (5) |

The propensity score, the probability that an individual receives the treatment. The propensity scores are then used to generate the counterfactual outcome for those in the treatment group, conditioning on observable characteristics.7

As noted in Caliendo and Kopeinig (2008) and Blundell and Costa Dias (2009), the MDID estimator has advantages over the augmented DID model. The standard DID approach assumes a linear relationship between treatment status and outcome. By contrast, the DID matching estimator assumes no functional relationship between treatment status and outcome, estimating this relationship nonparametrically. This departure from parametric assumptions increases the flexibility of the estimation procedure and improves the reliability of the estimates (Caliendo and Kopeinig, 2008). Although the DID matching estimator reduces the bias of the estimates, it also reduces the efficiency of the estimates. As a result, implementing this DID matching procedure leads to a reduction in significance of the coefficient estimates.

5. Data and descriptive analysis

The sample used in the present study matches information from three sources: course performance, the students’ transcripts, and the students’ university profiles. First, course performance data provide the outcome variables, the flipped/non-flipped indicator and attendance records. The outcome data is obtained from student assessments. We examine the effects of flipping for each of the three course units, as well as an overall comparison between the pre and post-tests. The assessment outcomes are expressed as a percentage of questions answered correctly. Second, with IRB approval, we match student performance data to their academic transcripts. This provides important controls for academic achievement, including the number of economics classes taken in previous semesters, a student’s cumulative college GPA, and the cumulative number of college credits earned in previous semesters. Finally, the student’s university profile provides background demographic information such as gender, race, and ethnicity, information on college major (both intended and declared), SAT college admission test scores, and whether a student lives on campus. The following variables are indicator variables taking a value of 1 if the answer is yes and 0 if the answer is no: respondent is white, respondent is black, respondent is male, on-campus resident, and business major.

Table 2 presents the descriptive statistics for the variables used, both for the pooled sample and for flipped and non-flipped classes separately. Looking at the means of the assessments, several patterns emerge. First, for both modes of course delivery, there was improvement between the pre-test and the post-test. Second, students in the flipped class showed greater improvement, on average, between the pre-test and post-test; however, for some pre/post comparisons, the increase
is quite modest. This increased improvement is most pronounced when looking at all 25 pretest questions compared with the 25 post-test questions. In this case, students in the flipped class showed an average improvement that was approximately 5% (1.35 questions) larger than their non-flipped counterparts. The flipped and non-flipped classes look similar with regards to the additional controls. With few exceptions, students in the flipped class have similar average values for the controls. The exceptions are as follows. First, students in the flipped class have taken slightly more economics classes on average. Second, students in the non-flipped classes have higher GPAs on average than their flipped counterpart, with a GPA of 2.92 (compared to a GPA of 2.71). Third, students in the non-flipped class have taken, on average, approximately six more credits (approximately two classes). Fourth, students in the flipped class have higher SAT scores, with an average that is 29 points higher than their non-flipped counterparts, approximately 0.25 standard deviations. While these differences are modest, we employ a matching approach to ensure that the treatment and control groups look similar over these measurable characteristics.

6. Results

Table 3 presents the results for the augmented DID regressions. Column (1) of Table 2 reports the results for the entire assessment, comparing gains between the pretest and posttest. These results suggest that between the two assessments, students, on average, improved their performance by 22% points. On average, students in the flipped class improved their performance by 6.1% more than the non-flipped classes.

Columns (2) through (4) of Table 3 report the results for the differences in scores on pretest and posttest questions divided into groups by unit. For both the Unit I questions and the Unit II questions there are no significant differential gains associated with flipping. By contrast, for the Unit III questions, students in the flipped section improved their assessment scores by a significant margin, improving by 11% points more, on average, than students in the non-flipped section.

Table 4 reports the results from the MDID estimation. While the MDID results follow a pattern similar to the augmented DID models, many of the effects are larger in magnitude. The most noteworthy of the results in Table 4 are the results of the pretest/post-test comparison of the entire assessment, reported in Column (1). On average, students answered 20% more questions correctly on the post-test than on the pretest. In addition, students in the flipped class showed significantly larger improvement between the pretest and posttest, improving by 7%, approximately 1.8 questions, more than students in the non-flipped classes on average. Further, for Units I and II, the results do not show a significant effect of flipping; however, the estimation for Unit III produces a significant effect of flipping. For the questions on the pre and post-test that came from Unit III, students in the flipped class improved 8.7 percent, approximately 0.7 questions, more than students in the non-flipped classes on average.

---

8 With one exception, the controls in the augmented DID models are insignificant. The exception is composite SAT scores, which have significant, positive relationship with assessment outcomes; a 100 point increase in SAT scores is associated with an increase in assessment performance of 2.4% points. These results are available upon request.
Table 4
MDID results.

<table>
<thead>
<tr>
<th></th>
<th>(1) All</th>
<th>(2) Unit 1</th>
<th>(3) Unit 2</th>
<th>(4) Unit 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flipped × post-test</td>
<td>7.036***</td>
<td>6.890***</td>
<td>5.504***</td>
<td>8.724***</td>
</tr>
<tr>
<td></td>
<td>(2.566)</td>
<td>(3.735)</td>
<td>(4.023)</td>
<td>(4.017)</td>
</tr>
<tr>
<td>Flipped</td>
<td>−3.256**</td>
<td>−2.28</td>
<td>1.024</td>
<td>−8.636***</td>
</tr>
<tr>
<td></td>
<td>(1.572)</td>
<td>(2.289)</td>
<td>(2.465)</td>
<td>(2.461)</td>
</tr>
<tr>
<td>Post-test</td>
<td>21.014***</td>
<td>11.104***</td>
<td>35.342***</td>
<td>17.833***</td>
</tr>
<tr>
<td></td>
<td>(1.819)</td>
<td>(2.648)</td>
<td>(2.852)</td>
<td>(2.848)</td>
</tr>
<tr>
<td>Constant</td>
<td>30.806***</td>
<td>34.277***</td>
<td>25.421***</td>
<td>32.287***</td>
</tr>
<tr>
<td></td>
<td>(1.112)</td>
<td>(1.618)</td>
<td>(1.743)</td>
<td>(1.740)</td>
</tr>
<tr>
<td>N</td>
<td>320</td>
<td>320</td>
<td>320</td>
<td>320</td>
</tr>
<tr>
<td>R-square</td>
<td>0.54</td>
<td>0.17</td>
<td>0.54</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Robust standard errors are provided in parentheses below each estimate. The dependent variable is the percentage of questions answered correctly. For column (1) the percentage is out of 25 questions. For column (2) the percentage is out of 9 questions. For columns (3) and (4), the percentage is out of 8 questions. The propensity score specification underlying the matching estimator includes the following controls: composite SAT scores, cumulative college GPA, cumulative college credits, number of economics classes taken, race ethnicity indicators, gender, on-campus resident status, program of study and the percentage of face-to-face classes the student attended.

* p < 0.05.
** p < 0.01.
*** p < 0.001.

7. Discussion

Our results indicate no negative effects of flipping and aggressively blending. In addition, our results suggest that there may be increased student performance that results from the flipped-blended format. Moreover, students in the flipped class showed more improvement in the third unit than in the first or second. We attribute this result to a student learning curve for successfully completing a flipped class, as the flipped format requires more self-discipline and independent work to prepare for the face-to-face sessions. Without such preparation, the face-to-face sessions are less valuable to students.

An analysis of these results coupled with additional observations helps to frame this study in the context of the extant literature. The magnitude of the effect from flipping can be better appreciated in comparison to other studies. While it is not a direct comparison, Chen and Lin (2012) find that the use of supplemental video lectures results in an average improvement of 4% on exams. Emerson and Taylor (2004) find that students in classes with interactive experiments improved their TUCE scores by 7–9% over the control group. In comparison, in this study the flipped class improved by 7% more than the non-

Table 5
Augmented DID-by quartiles of cumulative GPA.

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Quartiles of cumulative GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Q1</td>
</tr>
<tr>
<td>Flipped × post-test</td>
<td>6.151***</td>
<td>14.731***</td>
</tr>
<tr>
<td></td>
<td>(2.422)</td>
<td>(5.433)</td>
</tr>
<tr>
<td>Flipped</td>
<td>−1.374**</td>
<td>−5.327**</td>
</tr>
<tr>
<td></td>
<td>(1.421)</td>
<td>(3.245)</td>
</tr>
<tr>
<td></td>
<td>(1.71)</td>
<td>(4.587)</td>
</tr>
<tr>
<td>N</td>
<td>339</td>
<td>86</td>
</tr>
<tr>
<td>R-square</td>
<td>0.61</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Robust standard errors are provided in parentheses below each estimate. The dependent variable is the percentage of questions answered correctly on the entire assessment. All models include the following controls: composite SAT scores, cumulative college GPA, cumulative college credits, number of economics classes taken, race ethnicity indicators, gender, on-campus resident status, program of study and the percentage of face-to-face classes the student attended.

* p < 0.05.
** p < 0.01.
*** p < 0.001.
flipped classes (1.80 questions out of 25 questions). In addition, for the questions from Unit III, students in the flipped class improved 8.7% more than students in the non-flipped class (0.7 questions out of 8 questions).

Finally, similar to previous studies, we compare the flipping effect by student sub-groups. Figlio et al. (2013) find that Hispanic students, male students, and low-achieving (GPA lower than mean) students particularly benefitted from live lectures rather than video recordings and Emerson and Taylor (2004) find that low-achieving students and females particularly benefitted from the experimental curriculum. We do not find any significant differences by race or gender, but there is an interesting result for students with GPAs in the lowest quartile, summarized in Table 5. Table 5 reports the augmented DID regressions, dividing the sample by quartiles of cumulative GPA, used as a coarse measure of academic achievement in college. Students in the lowest quartile show large, significant average differential gains from being in a flipped class, improving their assessment scores by 14.7% points (approximately 3.67 questions) more than their non-flipped counterparts. It appears that the benefits from the face-to-face problem solving more than offset the costs of video lectures for the low-achieving students in our study.

8. Conclusion

The present study is, to our knowledge, the first controlled study of flipped-blended coursework in economics, pedagogical techniques gaining in popularity. The goal of combining the two pedagogies was to improve student learning while simultaneously reducing cost, by doubling the class size without requiring additional classroom space or additional faculty resources. The results of the present study suggest that flipping and blending the class may be a promising option for universities, reducing cost and the need for classroom space, without sacrificing student learning, and in some cases, enhancing it. The students in the flipped-blended class received 26% less face-to-face class time (5 fewer 75-min classes) with the instructor, but performed no worse and in some cases, performed better than the students in the modestly blended class.

The present study has its limitations. First, because this study uses a convenience sample, the estimated effects of flipping may not generalize to a larger population. As a result, the gains reported here might not appear in the population of undergraduate students. Second, it is possible that the benefits of flipping may occur neither in other economics courses nor in other subjects. Other courses or subjects may not lend themselves to the flipped format, and in those courses student learning would be lower because of the format. Third, it is possible that the estimated effect of flipped coursework may be due to other unmeasured individual characteristics. Yet, for this to be true, these characteristics must vary during the semester. Moreover, these characteristics are likely to be difficult, if not impossible to measure.

The aforementioned results and limitations suggest possible extensions to the current analysis. First, increasing the scale of the experiment over several universities and across the country would provide a more representative sample of college students, allowing the results to generalize to a larger population. Second, flipped coursework should be tested in a wide range of courses and subjects, to see where the gains lie. Even within economics, a careful examination of the effects of flipped coursework in other classes may provide further insights into the method’s benefits and drawbacks. Specifically, tool-based economics courses such as intermediate microeconomics may benefit most from this format by using class meetings to apply the theoretical tools to concrete applications.

Another direction for future research is the examination of differences in study behavior between students in flipped and non-flipped classes. Compared to students in the modestly blended classes, those in the flipped classes were required to be more prepared for their in-class meeting, both to complete the quizzes and to complete the in-class assignments. This change in behavior may result in better learning outcomes, and may spill over into other classes, as students adopt this preparation in their other courses. Thus, analyzing differences in behavior across class types may reveal exactly how the flipped format affects students. Despite the aforementioned limitations, the results presented here suggest that the flipped format is a promising pedagogical tool, warranting a careful analysis both of its effect on student learning and of its effect on student behavior.

Appendix A. Technical explanations

Technical explanations

Description of computation of difficulty and discrimination indices

To determine the level of difficulty of each question, we compute two commonly-cited statistics. First, we compute the difficulty index, which reports the percentage of each question’s responses that are correct. The difficulty indices for the twenty-five questions range from 0.08 to 0.73 on the pretest and 0.23 to 0.96 on the posttest.

Second, we compute the discrimination index (DI), which is computed as follows. We sort the observations by their overall score on the twenty-five questions. Then we count the number of students in the top 1/3 who answered the question correctly, and divide that number by the number of students in the top 1/3. The same value is computed for the lowest 1/3. Finally, we subtract the value for the bottom third from the value for the top third. Values of this index that are close to unity suggest the question is better at discriminating the high performing students with the low performing students. Likewise, values close to zero suggest the question does not discriminate well. For the twenty-five questions, the average
dissociation index ranges from \(-0.03\) to \(0.42\) on the pretest and \(0.02\) to \(0.57\) on the posttest. For more details of this procedure, refer to Kelley (1939).

Description of MDID estimation

We implement the MDID estimation using the STATA written code, “diff,” which uses the following specification (Villa, 2014). The MDID estimator requires the following assumptions in order to capture the true treatment effect. First, the MDID analog to the common trend assumption requires that

\[
E(Y_{it} - Y_{0it}|P, D = 1) = E(Y_{it} - Y_{0it}|P, D = 0)
\]

Here, the subscripts 0 and 1 denote, respectively the control and treated observations, and the subscripts \(t'\) and \(t\) denote, respectively the pre-treatment and post-treatment observations. The variable \(P\) denotes the estimated propensity score, and \(D\) as before is the treatment indicator. This assumption is less restrictive than the common trend assumption of the standard DID model.

Including the propensity score removes the assumption’s reliance on a specific parametric form, providing more flexibility in satisfying this assumption; however, using propensity scores requires additional assumptions. First, the “common support” assumption requires that the treatment (or non-treatment) is not deterministic, that each individual has a positive probability of being in either group. Letting \(Z\) denote the criteria that governs selection into treatment, this assumption can be written as \(0 < Pr(D = 1|Z) < 1\). In what follows, let \(S_0\) denote the set of observations that satisfy the common support assumption. Second, the “balancing property” must be satisfied. This assumption, sometimes called conditional independence assumption, requires that outcomes are orthogonal to treatment status, conditional on the selection criteria: \(Y \perp D | Z\).

Following the specification in Smith and Todd (2005), we estimate the effect of flipping as

\[
\alpha_{MDID} = \frac{1}{n_{1t}} \sum_{id=1}^{n_{1t}} \sum_{i,j} W(i,j) Y_{1it} - \frac{1}{n_{0t}} \sum_{id=1}^{n_{0t}} \sum_{i,j} W(i,j) Y_{0it} - \sum_{id=1}^{n_{1t}} \sum_{i,j} W(i,j) Y_{0it} - \sum_{id=1}^{n_{0t}} \sum_{i,j} W(i,j) Y_{0it}
\]

In the above expression, \(n_{1t}\) and \(n_{0t}\) denote the number of treatment and control groups in each time period, \(I_{1t}, I_{0t}, I_{1t'}, I_{0t'}\) denote the treatment and control groups in each time period, and \(W(i,j)\) are the weights assigned to each match between treatment and control observations.

References


