Eastern Economic Journal (2023) 49:433–456
https://doi.org/10.1057/s41302-023-00252-3

ORIGINAL ARTICLE

A Flipped Classroom Experiment in Growth Theory

Peter Mikek[1]

Published online: 18 May 2023
© EEA 2023

Abstract

This study utilizes a classroom experiment to determine the effectiveness of the flipped classroom in a growth theory module of intermediate macroeconomics. We compare the performance of small section flipped groups to control groups at a liberal arts college. The treatment groups watched growth lecture videos before class, while the control groups followed a traditional lecture format. The results provide no evidence of superior performance by the treatment sections. They show that the main determinant of a student's performance is high school quality. The careful preparation to minimize differences across students from different socio-economic backgrounds is essential.

Keywords Flipped classroom · Active learning · Teaching and learning economics · Class experiment · Growth theory

JEL Classification A22

Introduction

Flipped classroom pedagogy has attracted substantial attention for its ability to engage students and successfully bridge the gap between the teaching styles of professors and the learning styles of students (Lage, et al. 2000). We study the effectiveness of flipped classroom pedagogy on students' performance in a growth theory module in intermediate macroeconomics. Specifically, the research question is: Can the flipped classroom approach to a growth module in small sections of intermediate macroeconomics effectively improve students' performance on a high stakes test? Our contribution is two-fold: (a) we study the effectiveness of the teaching method within economic growth theory and (b) we investigate the method in an intermediate class with small sections at a liberal arts college.

✉ Peter Mikek
mikekp@wabash.edu

[1] Wabash College, Baxter Hall 37, 301 W. Wabash Avenue, Crawfordsville, IN 47933, USA

The reported benefits of flipped classroom pedagogy range from better short-term student learning outcomes and increased student-teacher interaction to more collaboration among students and individual pacing of learning. However, the challenges include stress for certain marginalized students working in groups, increased workloads, disparities in access to technology outside the classroom, and even the need for some adjustments in teaching and teachers themselves (Estes, et al. 2014; Akcayir and Akcayir 2018; Strelan, et al. 2020).

There are relatively few studies that have endeavored to rigorously measure the effect of a flipped classroom approach on performance in economics. The studies in flipped classroom learning specifically in economics have been mostly done for large introductory classes (Wozny, et al. 2018; Balaban et al. 2016; Craft and Linask 2019). Our approach is different, as we worked with small sections that, in general and given the academic culture of a small liberal arts college, require strong direct engagement of both students and the professor. Additionally, we focused on students in intermediate economics. Finally, we consider a growth theory module within a core economics class and endeavor to measure performance immediately following exposure.

Some authors acknowledge the challenges of teaching growth theory in undergraduate economics classes (Stein 2007). Additionally, our own experience from teaching the topic for many years indicates that students have substantial difficulties with the material. The attempts at overcoming these challenges suggest the use of active learning approaches (Swoboda and Feiler 2016), such as the flipped classroom. However, there is a lack of studies of its effectiveness (Purba et al. 2021), particularly for economic growth theory. Thus, we report on an attempt to rigorously measure the effectiveness of the flipped model in teaching intermediate growth theory.

The effectiveness of flipped pedagogy reported for economics varies substantially, from studies that find some evidence of improved performance (Lage et al. 2000; Wozny, et al. 2018) to others that find no improvement at all (Craft and Linask 2019; Setren et al. 2020). Thus, our motivation stems from these contradictory results and from a lack of studies on the effectiveness of active learning approaches in teaching growth theory in particular. We report on a classroom experiment comparing different sections of Intermediate Macroeconomics conducted through the module on growth theory. The experimental design followed the standard format of contrasting the results of treatment groups (flipped/problems) to control sections (standard lecture). In addition to the two sections in the same semester, we report the results for randomly chosen sections in different semesters (Craft and Linask 2019; Ficano 2019).

At the outset, we expected that the flipped classroom approach would improve students' performance on a test. However, some of our doubts about the approach stemmed from our experience with students' relatively short attention spans, a strong focus on grades vs. learning, and experience with previous "learn before the class" exercises.

The results clearly reject the null hypothesis of better performance for treatment groups on a brief test immediately following the experiment. Additionally, the main determinant of test success is the quality of high school, reflecting the socio-economic environment of individual students.

Related Flipped Pedagogy Literature

Strelan et al. (2020) perform a meta-analysis of the effects of the flipped model across many disciplines and report moderate improvements in learning. They suggest that the primary contribution of the approach is the opportunity for structured active learning. The benefit is contrasted with the substantial fixed cost of implementation (Vazquez and Chiang 2015) and time constraints (Mu and Paparas 2016; Vazquez and Chiang 2015; Roach 2014). Specifically for economics, Akcayir and Akcayir (2018), Estes, et al. (2014), and Purba et al. (2021) provide overviews of numerous flipped classroom studies and find modest effects on student learning. Lage, et al. (2000) claim that flipped pedagogy may address the mismatch in teaching and learning styles. Gulley and Jackson (2016) highlight the effective use of time and better retention. Several studies find improved preparation and student participation in class, including Balaban et al. (2016) and Hettler (2015). Boyle and Goffe (2018) expand beyond the flipped model to other active learning approaches. While the perception of benefits is not controversial (Purba et al. 2021), the rigorous measurement of the performance effects remains elusive. Roach (2014) suggests that “quantifying the gains from flipped learning in terms of increased test scores would be beneficial.” However, numerous authors limit their attention to the students’ perceptions of the approach (Roach 2014; Mu and Paparas 2016; Mendez-Carbajo and Malakar 2020).

There is much disagreement on the effects of the flipped approach on test performance. In economics, the research is almost exclusively done within large introductory economics classes (Wozny, et al. 2018; Caviglia-Harris 2016; Yamarik 2019; Balaban et al. 2016; Setren et al. 2020; Calimeris and Sauer 2015). Very few authors have implemented the flipped model in intermediate classes (Gulley and Jackson 2016; Yamarik 2019; Webb et al. 2021), and we found no study that would implement it in an intermediate class focused on growth theory. The available literature suggests a range of findings about its effectiveness, from a positive effect to no effect whatsoever.

Flipping a classroom in a large introductory econometrics class, Wozny et al. (2018) find improvement on a high-stakes test, with a similar long-term effect. Similarly, Balaban et al. (2016) also find improvement on the final exam for sections of Principles of Economics in which a flipped approach was used. They report that the magnitude depends on the particular learning objectives. Yamarik (2019) conjectures that the flipped format may increase learning productivity, with students earning similar grades with worse attendance. Caviglia-Harris (2016) compares the traditional lecture approach to complemented lectures and an entirely flipped classroom approach, and reports improvement for the flipped approach. Interestingly, Cosgrove and Olitsky (2020) report modest gains from the approach and expand it with a research-based learning strategy for which students are expected to do some independent research before class discussions, and here they find stronger effects.

Unlike this paper, almost all of these studies measured performance with some delay, on the final or midterm exams

Considering partially flipped classes, studies examined completely flipping only a part of the course or some aspect of the course for the whole semester. Early studies of Chen and Lin (2012) and Flores and Savage (2007) report better performance for students in classes that supplement their recorded lectures. Lombardini et al. (2018) and Singh (2020) adjust some aspects of their classes for the whole semester. For example, Lombardini et al. provide their own lectures outside class but still lecture in class on elements that students asked about and combine this with in-class group activities. They both report an increase in the likelihood that students will pass the class. Similarly, Caviglia-Harris (2016) finds that the partially flipped classroom that had "mini" lectures complemented with online videos had a positive effect but was less effective than the full flip. Finally, Wozny et al. (2018) selected 10 out of 25 lessons for a flipped treatment and found that students did better on the questions for those lessons on their final exam.

In more nuanced results, Calimeris and Sauer (2015) report an initial short-term "negative adjustment period" to a flipped classroom approach, after which students improved their performance. However, Craft and Linask (2019) and Setren et al. (2020) find no evidence of improved learning outcomes in economics at the end of a semester. While working with a different active learning mode (simulations), Green (2014) reports that treatment groups actually performed worse than the control group. This raises additional questions about the implementation of active learning strategies and a flipped classroom approach in particular.

Highlighting the importance of careful preparation in implementing flipped pedagogy, Webb et al. (2021) suggest an interesting distinction. They separate the class preparation into "didactic" and "non-didactic" pre-class material. They first mimic the material and format of standard lectures and found that it is the type of flipping that matters. In particular, they report a positive effect when the flipping does not mimic the traditional lectures, but there is no effect when they do.

Finally, Ficano (2019) reminds us that teaching and learning is not done in a socio-economic vacuum. She reports that students with more developed math literacy and majority students benefit from a flipped pedagogy approach. In contrast, however, the pedagogy has a negative impact on weaker students and on minority students. Similarly, Setren et al. (2020) find that the flipped approach persistently widens the achievement gap, with most benefits accruing for white, male, and higher achieving students.

Teaching Undergraduate Growth Theory

The literature on teaching undergraduate economic growth theory is scarce even though growth theory is a fundamental topic in undergraduate economics. However, due to a number of challenges, instructors frequently choose to skip it entirely or treat it only marginally (Acemoglu 2013).

There are a number of challenges to teaching undergraduate growth theory. The first among the challenges is the difficulty students have with understanding the mathematical structure and manipulation of variables in growth models (Ballard and Johnson 2004; Bosshardt and Manage 2011). Similarly, students need a sufficient foundation in microeconomics, macroeconomics (Stein 2007; Benge and Wells 2002) and statistics (Elmslie and Tebaldi 2010) in order to understand growth theory.

Another challenge is incorporating enough empirical evidence to bolster students' conceptual understanding. This imposes some fixed costs (Vazquez and Chiang 2015) and requires additional class time (Mu and Paparas 2016; Vazquez and Chiang 2015). Elmslie and Tebaldi (2010) propose data analysis exploring the effect of corruption on GDP for better understanding of causation and the difference between GDP level and long-run growth rate effects. Barreto (2016) provides an example of how the Maddison historical data on GDP (Bolt and vanZanden 2020) easily fits this purpose.

Furthermore, developing an understanding of the conceptual difference between proximate and fundamental causes of economic growth (Acemoglu 2013) requires additional effort to connect them as they are not "readily amenable" to modeling (Mankiw 2019). This requires some insights into the role of institutions, culture, geography, and political factors in the dynamics of the proximate causes of growth.

Students that are used to the equilibrium approach may find the dynamics in growth models confusing (Stein 2007). Despite this, Acemoglu (2013) points out that the use of aggregates in macroeconomics is a "huge methodological change" for students, as it departs from what they are used to in microeconomics. Therefore, he suggests including more growth theory in the undergraduate curriculum. He argues that the methodological switch does not apply to growth theory since it is based on "tools and insights on which economists agree and do not require an entirely different mindset than basic micro" (Acemoglu 2013). He further suggests starting growth theory with the following aggregate production function:

$$Y = F(K, H) \tag{1}$$

where Y is GDP , K is the physical capital stock, and H is the efficiency units of labor. Using H allows for quickly extending expositions to human capital, technology, and efficiency. A similar approach is used in some standard textbooks (Mankiw 2022). It substantially simplifies the dynamic aspects of the model to a simple steady state relationship:

$$sf(k) = (d + n + g)k \tag{2}$$

where parameters are given exogenously: s—saving rate, d—depreciation rate, n—population growth rate, and g—growth rate of GDP due to improved total factor productivity. $f(k)$ represents GDP and k physical capital, both expressed in per effective worker terms. Thus, the left-hand side gives saving and the right-hand side the replacement capital or break-even investment. Similarly, the graphical representation is rather simple and allows for easy handling of changes in proximate causes (parameters) of growth.

However, as Fig. 1 illustrates, an increase in the growth rate of total factor productivity (g) expands the number of "effective workers" and thus lowers steady state

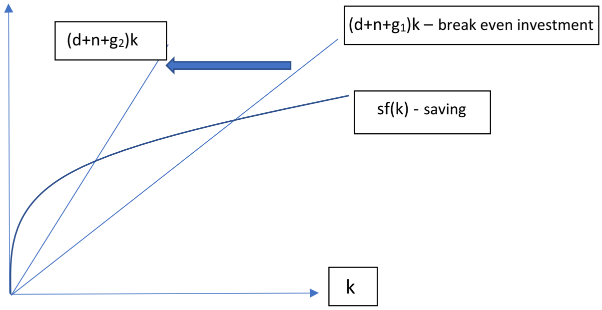

Fig. 1 Change of steady state due to faster technological improvements in the Solow growth model

capital and output per effective worker. At the same time, this implies improvements in living standard (output per worker) since more output is produced. Students find this confusing. Therefore, the framework presents another conceptual difficulty: understanding variables expressed in "per effective worker" terms.

These challenges invite active learning approaches (Swoboda and Feiler 2016) to the teaching of undergraduate growth theory. For this purpose, authors have suggested a variety of teaching techniques, such as problem sets, growth scenarios, in class discussions (Hettler 2015), and data analysis exercises (Elmslie and Tebaldi 2010). In summary, the desire to improve the teaching of growth theory ultimately led to the present experiment and to asking whether the approach actually works for small groups at the intermediate level.

A Model of Student Performance

Consider the performance production function proposed by Balaban, et al. (2016) describing the transformation of student endowments, instructional capital, and students' effort into education outcomes, such as performance on an exam. The endowment of an individual student i is captured by a vector A_i and includes both a student's human capital embodied in educational endowment due to experience and past effort, and other characteristics of an individual student. Each student supplies the learning effort, denoted by E_i. The instructional capital that is common to all the students in a class is denoted by a vector K that includes characteristics of the professor, course and classroom, and teaching methods. Let Y_i denote the educational outcome for student i. Then, the production function for an individual student can be given as:

$$Y = f(K, E, A] = K^a E^b A^c \tag{3}$$

where a, b and c measure responsiveness of the performance outcome to changes in instructional capital, effort supplied, and a student's endowment, respectively. To study the effects of a flipped classroom approach on exam performance, this non-linear production function can be approximated in the following manner:

$$Y_i = g\left(F_i, A_i, X_i, E_i, K\right) + \varepsilon_i \tag{4}$$

where ε_i captures the impact of unobservable characteristics on a student's performance. F_i is a flipped classroom indicator variable. Data across several semesters require an experimental design seeking minimal variation in the common instructional capital in vector K. Exogenous characteristics of the students are captured by the vector X_i, including race and high school quality. The measures of ability incorporated in the vector A_i are captured before enrollment into college.

The likely correlation between unobservables that explain ability and performance may lead to a biased estimate of the marginal effect of ability on performance. However, Balaban, et al. (2016) point out that this "is not what we seek to measure." Instead, our main focus is the marginal effect of the flipped classroom format on performance. As the choice of the flipped classroom by a professor is exogenous to individual students and their ability, the marginal effect is not plagued by bias as ability and class format are not correlated.

Nevertheless, Balaban, et al. (2016) suggest that the choice of the form of instruction may affect a student's effort during the semester. In this case, the total marginal effect of the class format on performance should include the indirect effect of effort on performance. But they also point out that "it is difficult to find exogenous individual characteristics that impact effort but do not affect performance" and therefore focus on the total effect not conditional on effort.

Methodology and Data

Experimental Design

The experimental design first included two sections of the same course in the same semester in 2016. Students were quasi randomly assigned to the treatment and control sections as they chose them in advance not knowing that they might be in one of the treatment groups at the time of registration. To eliminate the self-selection bias, they were invited to participate in the study only after the course had already started. They could not choose the instructional format. The only choice they had was whether to allow the investigator to use their data or not. All students with the exception of two elected to allow their data to be used in this study. If a student decided not to give consent, they were still part of the same instructional format for the given semester, but their data was not included in the study. None of the students dropped the class before the completion of the module and none was a transfer student. The course is required for all economics majors, further limiting possible self-selection bias. It is a part of the core for the economics curriculum (along with Microeconomics and Econometrics) and is needed for normal work in advanced classes. All students were economics majors who had met the pre-requisites of Calculus and Principles of Economics, and all had seen some rudimentary treatment of growth in Principles.

Seeking the minimal variation in instructional capital (K), the classes met in the same classroom, had the same homework, and were taught by the same instructor, eliminating possible effects of teaching style. No alternative section or instructor was available for this course during the time of the experiment. The sections were taught one after another in midmorning first semester and at the same time across the semesters. After the initial phase, we followed a standard procedure (Craft and Linask 2019; Ficano 2019; Balaban et al. 2016) in collecting additional data for a single section over several randomly chosen semesters. The semesters for the flipped model were chosen randomly by flipping a coin. Therefore, the only variation on the initial experimental design is that instead of directly comparing two particular sections within a semester, we compared the results for sections across several semesters (6 semesters, two of which were treatment groups). All sections were small and ranged between 8 and 32 students.

In contrast to most studies from large universities (Balaban et al. 2016; Vazquez and Chiang 2015; Mu and Paparas 2016; Roach 2014), the experiment was carried out at a small liberal arts college. Small sections and the institutional culture provide a very dynamic, student-centered, and interactive teaching environment with plenty of direct involvement on the part of students. While the improvements in in-class engagement in flipped classes seem to be substantial for large classes at large universities (Balaban et al. 2016), they are not the sole benefit of the approach. In particular, students in flipped classrooms can benefit from a more personalized learning experience, deeper understanding of the material, and gaining agency in the self-directed learning with some guidance but no hand-holding. The approach can better accommodate different learning needs, styles, and preferences, and corresponding teaching styles and preferences. It develops independent learning skills, as students can progress at their own pace and return to the material several times. The class time can be used for fine-tuning of the acquired understanding. Therefore, we expected to find measurable effects of the intervention even within an already highly interactive teaching environment.

Overall, students actively participated in each class, absences were extremely rare, no one dropped the classes, students had regular access to their professor, and the whole course was set up so as to stimulate students' engagement and discussion. Along with regular homework, this provided a consistent level of student effort across different sections.

The experiment was a novel approach for our students in a number of ways. They had not been exposed to a flipped classroom before, at least not at the college. Additionally, some forms of active learning are rather different from the flipped approach. In particular, some active learning forms actually allow for different levels of engagement for students. For example, group work that is not well structured can allow some students to "hide behind" more active group members, which doesn't work in a flipped classroom. Similarly, many other forms of active learning are focused on work in class, such as discussion-based classes or group work discussions. Frequently, such discussions invite a variety of different interpretations, where students exercise their creative imagination, and do not include rather narrowly focused, and to some extent technical, problem sets. While both have benefits, they are different in nature. Moreover, the flipped classroom approach requires that each student take some control of the learning process. However, taking such control is a slow and uneven process of maturation. Most students are somewhere in the middle

Table 1 Activities for treatment and control groups

	Prior to class	In class	After class
Flipped classroom (treatment)	Readings Watching video lectures A few questions about the videos	Problem examples developed on board Collaborative work on problems in pairs Extensive Q&A session	Problem set
Standard lecture (control)	Readings	Standard lecture with problem examples Q&A session	Problem set

of it when they come to this class, beyond initial college level classes, but not yet fully mature. Thus, the approach was different enough from other forms of active learning at the institution to be considered a valid treatment.

Description of the Study

The control group followed the standard lectures combined with in-class problems and homework. The treatment group, however, watched growth lecture videos online before coming to class and used the class time for discussion of what they had learned and for solving problems, similar to those later used on the test (see Appendix 2). In class conversations, the professor used some leading questions but encouraged students to pose their own questions, as they are used to small group discussion-based classes within the liberal arts curriculum. Each class involved work on problems, by the professor and in pairs, followed by homework. Table 1 summarizes the learning activities for treatment and control groups. The experiment was applied just for the specific growth theory module part of the course. In an environment where students are used to participating in class, the adjustment to a different format was relatively easy.

Unlike almost all studies reviewed, we opted for screencasts not prepared by the instructor. The publicly available screencasts were recorded by H. Barreto at DePauw University (Barreto 2016). They are prepared for teaching with heavy use of Microsoft Excel. While some of the screencasts deal with empirical data (Madison's international historical data on GDP (Bolt and vanZanden (2020)), the main objective of the screencasts is to overcome the mathematical problems with a hands-on modeling approach using concrete numerical information and corresponding graphs (Barreto 2015). The students can change the parameters in the Excel files that accompany the screencasts and immediately see the results of these changes directly in the graphs representing the models. Thus, there are two main goals for these screencasts: (1) to explain the main model characteristics with the use of concrete numerical information, which avoids dealing directly with complicated math; and (2) to provide students with a hands-on lab-type experience for exploration of changes in parameters on the model outcomes in graphical and numerical form. The use of Excel is not an obstacle for these students, as all of them had substantial experience with Excel and with the particular way of using it in the classroom from their previous economics classes. We assigned only those screencasts that closely aligned with the textbook (Mankiw 2022). Students watched the screencasts covering a number of growth topics, such as finding initial steady state in the Solow model, finding the golden rule steady state, extensions to population growth and technology improvements, some comparative statistics, and some transitional dynamics of the model.

While watching the screencasts, students had an opportunity to work along with the presenter, changing the parameters in the Excel file themselves. Some of the screencasts demonstrate how to find a steady state or how to calculate the golden rule steady state. The screencasts directly address questions that were nearly identical to those given on the test (provided in Appendix 2), including one of the tables included in the test (question 3). After watching the videos, in-class time for the treatment group was devoted to solving problems very similar to those given on the test, both by the professor, in pairs, or even individually. The classes were relaxed in atmosphere and like workshops, with many questions about the problems and material. After the class, students were assigned homework with similar problems again.

Most previous studies reported using multiple choice and essay questions after some delay on final exams (Caviglia-Harris 2016; Lombardini et al. 2018). However, the students here were given the instrument measuring performance immediately after exposure to the module, during the class session after the conclusion of the module. The high-stakes test was worth 15% of their final grade. We would highlight that they were asked standard questions about growth theory, such as to find the initial and golden rule steady state for the Solow growth model, draw a canonical Solow graph to illustrate a change in a model's parameter, find initial saving rate, and fill in missing values in the table that shows steady states for different saving rates. In short, these are standard questions that quite closely follow the exposition of the material in the textbook, screencasts, in-class problems, and homework problem sets.

Data

We collected data on students' ability (A), including high school background, such as GPA (Yamarik 2019; Caviglia-Harris 2016; Roach 2014)[1], and standardized tests, SAT/ACT (Freeman et al. 2011; Haak et al. 2011). As we frequently encounter students with inadequate mathematics skills (Bosshardt and Manage 2011), we included information about their initial placement in mathematics (remedial or calculus). The students' individual characteristics (X) include some demographic indicators, high school rating, and dummies for being first generation attending college and Pell grant recipient. Most of the data on high school and demographics is collected by the college as part of the admissions process. It was matched with test scores by name based on the signed permission of students. The initial semester with two sections was Fall 2016, and we completed collection of data in 2019, prior

---

[1] While GPA is not a standardized measure, we include it for comparison with previous studies.

Table 2 Summary statistics for the main variables

	Total		Standard		Flipped	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Test score	68.8	19.56	70.1	19	63.4	18
Flip	0.25	0.44	–	–	–	–
SAT total	1630	262	1668	256	1516	252
SAT math	638	84	636	83	646	91
High School GPA	3.8	0.5	3.7	0.5	3.9	0.7
Math placement	0.25	0.44	0.25	0.43	0.26	0.45
First generation	0.22	0.41	0.14	0.35	0.43	0.51
Pell grant recipient	0.20	0.40	0.17	0.38	0.26	0.50
High school rating	0.6	0.21	0.6	0.2	0.6	0.2
No. of observations	92	-	69	-	23	-

to the covid-19 pandemic. Table 2 provides summary statistics for all sections and separately for the control (standard approach) and flipped model.

While our sample size is small, it is relatively close to some other studies that focused on small sections (Caviglia-Harris 2016; Wozny, et al. 2018). Some studies find gender significant (Setren et al. 2020) and others report that "gender had no statistically significant impact on learning outcomes" (Lombardini et al. 2018). Since the experiment was carried out at a college with only male students, it is, unfortunately, not possible to account for gender variation. Table 2 reveals that about a quarter of the students were in a treatment section, and the same share completed a remedial math class before the course. About one-fifth of the students in the sample were first generation and Pell grant recipients, with a bit higher share of these in the flipped model. The groups were comparable in average high school rating and in math SAT scores. Both groups had almost the same share of low math placements. However, the control group had a bit higher SAT total score and a bit lower high school GPA. The mean test performance was only about 69% on average across all sections, but was higher in the standard than in the flipped group. Figure 2 below indicates that the difference between the means is not due to outliers.

With almost the same standard deviation, the standard group has a larger proportion of students earning high grades (above 80). In contrast, only a small portion of students is in that range for the flipped group. The flipped test scores are most commonly in the range of 60–80.

While the other variables are standard measures of ability, the high school rating merits some explanation. It is a number between 0 and 1 and is determined from a variety of quality indicators, such as student-teacher ratios. The mean score for students at this college[2] over the last 10 years is around 0.6 and

---

[2] We are grateful to D. Dalenberg for assistance with the initial raw data and explanation of high school rating.

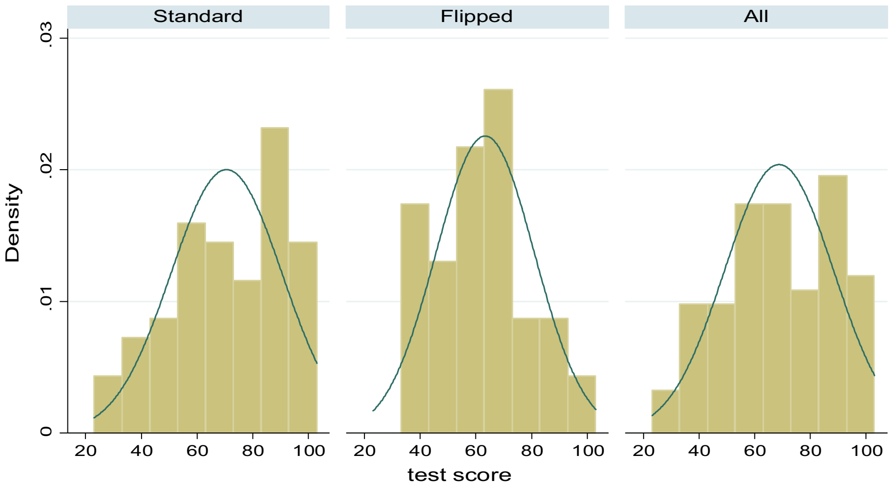

Fig. 2 Distribution of test scores for standard, flipped and all sections

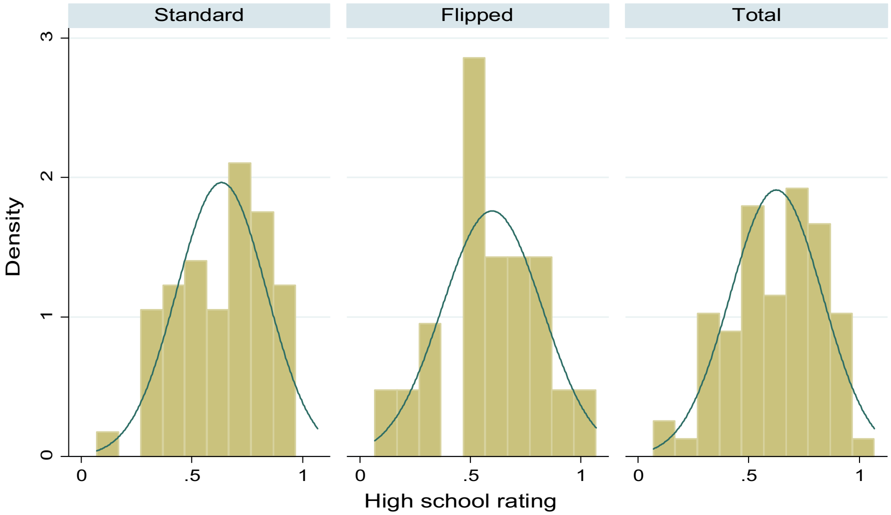

Fig. 3 High school rating

for the particular sample used here 0.625. Figure 3 shows its distributions. With almost the same average and standard deviation, the groups do not differ substantially with the exception of some concentration around 0.5 for the treatment group.

Table 3 Flipped average mean test score is lower

Group	Mean	SE	No. obs.
Control	70.59	2.40	69
Treatment (Flip)	63.36	3.67	23

Empirical Model

The cross-sectional data was analyzed through standard statistical tests and regression with robust standard errors so as to take into account possible changes in variability across sections, and tested for heteroscedasticity. We estimated the following equation:

$$\text{testscore}_i = \beta_0 + \beta_1 \text{flip}_i + \beta_2 A_i + \beta_3 X_i + \beta_4 \text{IT}_i + \varepsilon_i \tag{5}$$

testscore is the test score in percentage points, *flip* is a dummy variable for the treatment sections, the vector of ability controls $\boldsymbol{A}_i$ includes measures of standard test performance (SAT total and SAT math), high school performance (high school GPA), and math placement. The vector of individual characteristics ***Xi*** includes high school quality (high school rating), dummies for first generation and for recipients of Pell grants, and some demographic characteristics. We complemented the equation with some interaction terms (***IT***). The model was estimated using robust standard errors clustered by sections.

Following Balaban et al. (2016), we exclude effort from this specification. They state, "It is important here to not include individual effort as an explanatory variable because effort during the semester may be influenced by the instructional format." Additionally, small sections with almost perfect attendance, completion of homework and very active participation limit variations in effort between groups. Finally, it is very difficult to find a reliable measure of effort, which may be the reason that most studies don't include it (Setren et al. 2020; Balaban et al. 2016; Wozny, et al. 2018).

Results

The Same Semester

As the sample size for the first semester with two sections was very small (24), we only summarize the findings here. Surprisingly, the treatment group on average scored about 7 percentage points lower than the control group (75 vs. 82 with SE 3.02 and 4.08 respectively). We rejected the null hypothesis of higher performance for flipped model at 10% significance level. Controlling for a number of determinants for ability (A) and personal characteristics (X) renders the coefficient for treatment dummy (flip) not significant (p-value 0.428). The preliminary result is in stark contrast to our expectations as we find no evidence of improvements on a standard high stakes test for the flipped model.

Beyond the Same Semester

The results for the initial hypothesis for data across several semesters are provided in Table 3 below.

We found that the mean test score for the treatment group was about 7 percentage points lower. There is sufficient evidence to reject the null that the treatment has a higher mean at the 10% significance level.

Expanding beyond the basic comparison of means, we controlled for a number of variables capturing the students' ability (high school academic success, their performance on standardized tests, the dummy for the initial math placement), personal characteristics (high school rating), the dummy for the change in SAT (SAT changed its structure during the period of the study), and a number of interaction terms for the flipped sections. The results are shown in Table 4 below.

It is striking that, in general, we did not find any significant variables, with two exceptions. Neither initial placement in math classes (remedial or usual college level calculus) nor high school GPA appears as a significant determinant of the test scores in this experiment. Given our experience over the years with students struggling with mathematics in economics courses (Ballard and Johnson 2004), this is rather surprising.

For ability, high achieving students (determined by SAT score) in flipped sections did a bit better on the test by about 2 percentage points for those with SAT scores higher by 50 points. However, across all specifications where included, it appears that the students who graduated from high schools with higher ratings did substantially better. Students with a high school rating higher by 0.1 (high school rating is between 0 and 1) on average scored between about 2 to 3 percentage points higher. Including controls for first generation and Pell grant recipients doesn't change the main results, similar to Caviglia-Harris (2016).

Here are some further details on high school rating as it appears to be of central interest. As mentioned above, the mean for our sample is 0.625, with almost the same mean and same standard deviation for both groups. The distributions of high school rating for each section are provided in Figure 3 in Appendix 1. While there is clearly some variation between different sections, there is no evidence to believe that the treatment group (sections "We are grateful to D. Dalenberg for assistance with the initial raw data and explanation of high school rating." and 6) is substantially different from the control group (Sect. 1, 3, 4, and 5—see data section above). Based on data from the US Department of Education, National Center for Educational Statistics, State Department of Education websites, and non-profit organizations, the ranking system uses a number of factors, including:

- The percentage of students taking Advanced Placement courses
- Number of Advanced Placement courses offered at the high school
- The percentage of students taking honors courses

Table 4 Estimates for test performance: importance of high school quality

Dependent: test score	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Flip	−1.8917	−53.8848	−12.5568	−24.5652	20.4043	−30.1137	−36.0587
	(0.741)	(0.144)	(0.765)	(0.559)	(0.126)	(0.528)	(0.507)
High school rating	19.5589*	27.6955**	29.0069*		28.2284*	24.1023*	25.0201*
	(0.087)	(0.038)	(0.065)		(0.071)	(0.055)	(0.063)
SAT total	−0.0038	−0.0104	−0.0106	−0.0003		−0.0101	−0.0102
	(0.778)	(0.444)	(0.463)	(0.978)		(0.490)	(0.498)
SAT math					−0.178		
					(0.652)		
SAT change dummy	−7.9017	−2.3156	−1.7205	6.8205	−2.353	−1.6052	−1.3256
	(0.375)	(0.799)	(0.816)	(0.303)	(0.731)	(0.837)	(0.868)
Math Placement	−5.5279	−4.5095	−3.6391	−3.1160	−4.500	−5.2445	−5.0359
	(0.387)	(0.476)	(0.576)	(0.587)	(0.514)	(0.416)	(0.446)
High school GPA			3.8797	0.6284	3.3063	2.2207	2.1744
			(0.470)	(0.895)	(0.568)	(0.665)	(0.685)
First generation						8.0926	8.5912
						(0.349)	(0.346)
Pell grant recipient							−0.8986
							(0.912)
Flip*SAT total		0.0427*	0.0404**	0.0388**		0.0386**	0.0423*
		(0.053)	(0.024)	(0.022)		(0.018)	(0.051)
Flip*SAT math					0.0094		
					(0.857)		
Flip*High school rating		−26.1636	−27.2457		−24.3358		
		(0.269)	(0.164)		(0.247)		
Flip*High School GPA			−9.7434	−11.3680	−14.7087*	−8.7780	−8.7821
			(0.210)	(0.138)	(0.096)	(0.338)	(0.340)
Flip*First generation						−5.0418	−6.9145
						(0.632)	(0.546)
Flip*Pell recipient							4.4525
							(0.746)
Constant	66.2409***	71.1124***	55.6305*	68.4632***	52.2417	63.4483**	63.3271*
	(0.005)	(0.004)	(0.090)	(0.010)	(0.181)	(0.035)	(0.053)
Observations	78	78	78	91	78	78	78
R^2	0.065	0.126	0.135	0.095	0.095	0.132	0.133
AIC	602	692	695	805	698	697	701

p-values in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

- The average scores achieved by students on standardized tests
- The percent of students going on to college
- Teacher-student ratios

Table 5 Estimates for the final specifications: better performance for better high school quality

Dependent: test scores	(1)	(2)
High school rating	29.6225*	26.7072*
	(0.065)	(0.077)
Flip	−51.3020*	19.3407
	(0.057)	(0.548)
Flip*high school rating	−28.2740	−25.5626
	(0.126)	(0.211)
SAT total	−0.0065	
	(0.566)	
Flip*SAT total	0.0417**	
	(0.019)	
SAT math		−0.0048
		(0.897)
Flips*SAT math		−0.0123
		(0.807)
Constant	63.0797***	55.9949**
	(0.001)	(0.018)
Observations	78	78
R^2	0.119	0.067
AIC	688.6	693.1
H_0*: constant variance:chi*^{2}*(1)*	1.01	0.44
pval	(0.314)	(0.51)

p-values in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

These factors are an indication of the resources available to students at a high school. Therefore, it is the quality of the educational environment that makes the difference in student performance on the test. The students that went through a higher quality of academic preparation in high school are better prepared to absorb information in a different, non-standard way, such as flipped-classroom pedagogy. The resources available to high schools heavily depend on the socio-economic environment of the schools and therefore their students.

Including a number of covariates determining a student's academic performance in high school introduces some level of multicollinearity. This, in turn, expands the variance for the estimates and may render them not significant. Thus, we pruned some of the regressors and in Table 5 present the final two specifications.

H_0 of constant variance cannot be rejected, indicating no bias due to heteroscedasticity. The results indicate that both ability (SAT) and personal characteristics (high school rating) remain important. The high school rating improves the test score by about 3 percentage points for a 0.1 increase in the index (1.3 percentage points for flipped). Similarly, the SAT interaction term is a significant predictor of test score. Students in the treatment group scored on average about 2 percentage points higher for each additional 50 points on the SAT.

Table 6 Estimates including race

Dependent: test score	(1)	(2)	(3)	(4)	(5)	(6)
High school rating	28.7180*	27.5280*	27.7332*			
	(0.056)	(0.068)	(0.077)			
Flip	−53.7389*	−51.0115*	−51.2151*	−61.0684**	−61.0910**	−60.9921**
	(0.053)	(0.060)	(0.059)	(0.041)	(0.041)	(0.041)
Flip*High school rating	−26.5514	−28.7602	−28.3847			
	(0.152)	(0.125)	(0.138)			
SAT total	−0.0065	−0.0063	−0.0065	−0.0020	−0.0020	−0.0015
	(0.566)	(0.581)	(0.569)	(0.825)	(0.831)	(0.867)
Flip*SAT total	0.0424**	0.0417**	0.0417**	0.3550*	0.0356*	0.0350
	(0.019)	(0.021)	(0.020)	(0.061)	(0.060)	(0.064)
Non-White*	3.9231			1.9681		
	(0.543)			(0.678)		
Asian		2.8779			2.7910	
		(0.719)			(0.635)	
African–American			0.5728			−7.0912
			(0.961)			(0.476)
Constant	62.0545***	62.7791***	62.9928***	73.4167***	73.3606***	73.5563***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Observations	78	78	78	92	92	92
R^2	0.123	0.120	0.119	0.073	0.0736	0.0767
AIC	690.2	690.6	690.6	810.2	810.1	809.8
H_0:*constant varian.:chi*^{2}*(1)***	1.10	1.13	1.05	0.06	0.04	0.11
pval	0.295	0.288	0.306	0.807	0.848	0.737

p-values in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

*Non-white includes African–American, Asian, Native Indian, and Hispanic students

**Breusch-Pagan/Cook Weisberg test for heteroscedasticity

Additionally, we find that the coefficient on flip dummy is significant at the 10% significance level in one of the specifications. What is unexpected, however, is that the sign is negative. This means that the flipped classroom experiment actually decreased the performance for students on the test on average. This underscores the need to prepare students very well for such an approach. While the basic idea is that the students absorb some material in advance and then expand upon it, the proper preparation is crucial (Estes, et al. 2014).

Finally, we examined whether race helps predict performance on the test. The results are given in Table 6 below. 25% of the students in the sample were non-white (including Hispanic), 5% were African-American, and 14% Asian. Race doesn't appear to be an additional direct predictor of the performance on this particular test, similar to the result of Caviglia-Harrris (2016). However, there is a high correlation between race and SAT achievement (Card and Rothstein 2007). Similarly, there is likely a correlation between race and high school quality. In addition, different racial groups have, on average, very different experiences in their high school environments. We suspect that the worse general socio-economic conditions faced by some racial groups are picked up by both flip*SAT interaction terms and the high school rating. Thus, these two variables already partially reflect the worse conditions experienced by some racial groups. The correlation between test score and high school rating is 0.567 for African-American students, while it is only 0.222 for white students. However, the sample size is small and race is likely correlated to a number of other variables not included here, so it merits exploration in future research.

Discussion

The results suggest the crucial role that high school quality plays in students' success in general, and for benefiting from the flipped classroom pedagogy in particular. The differences in high school rating reflect resources available at high schools. Thus, similarly to Ficano (2019) and Setren et al. (2020), we find that students from more affluent environments are more likely to benefit from a variety of available resources in their high schools. Of particular value is close work with teachers stimulating higher order thinking, such as concept transfer and evaluation, in more resource-rich high schools. This prepares students to be able to take advantage of different teaching approaches. It is possible that some of them have already been exposed to a broader variety of pedagogical approaches.

In line with Ficano (2019) and Setren et al. (2020), we find that starting with a level playing field for all students requires appropriate and thoughtful preparation for students, particularly those from weaker socio-economic backgrounds. The preparation should include conversation with students about the topic in advance of the flipped classroom exposure, preparation for some unknown vocabulary with basic field terminology they are likely to encounter, assigning some leading questions, assigning some preparatory readings, etc. This is to set up a common point of departure for all students. Furthermore, resource-rich schools have means to foster individual exploration in students relatively early, which stimulates intrinsic motivation. Students with strong intrinsic motivation are most likely to be able to take advantage of the approach.

Going into a flipped classroom arrangement with an unexamined set of assumptions about students' knowledge and backgrounds will disadvantage minorities and marginalized groups (Ficano 2019). Thus, the teacher should think carefully about the assumptions about students' learning and their starting point. For example, some students from less fortunate socio-economic backgrounds may have particular trouble transferring information into a broader context or connecting videos they watch before class to the textbook. Additionally, the choice of "non-didactic" pre-class materials may be crucial (Webb et al. 2021). We are not suggesting that the videos used in this experiment are not a useful pedagogical tool. Just the opposite, as we have used them very successfully over the years as both a supplement to regular lectures, as part of homework assignments, and as a model for teaching some classes.

However, Webb et al. (Webb, et al. 2021) caution us to be careful in our selection of pre-class exposure materials given the initial knowledge (in a very broad sense) of our students.

For highly heterogenous groups of students, the approach may be better suited for more advanced classes where students have had some exposure to the field and have had an opportunity to develop some level of interest in the topic. Additionally, students in intermediate classes are more likely to have better synchronized starting points. Furthermore, in colleges with small sections, the teachers have an opportunity to get to know students better. Thus, using the flipped approach in large introductory classes without very careful preparation may be pedagogically counterproductive.

For this particular sample and in stark contrast to large universities, the approach probably added relatively little to the already highly interactive in-class environment in which students are accustomed to collaborating with their peers. However, beyond in-class participation, the benefits may be of a more long-term nature and may not be captured clearly in the short run, as found by Calimeris and Sauer (2015). Finally, as Strelan et al. (2020) point out, the main benefit of the approach is likely due to the opportunity for "structured, active learning and problem solving."

While the findings here align with those in the literature (Green 2014; Craft and Linask 2019; Ficano 2019; Setren et al. 2020), there are several caveats concerning the research design. First, while close to other similar studies, the sample is relatively small, which is due to the fact that the study was done at a small liberal arts college with small sections. Second, the results cannot account for possible gender variations as there were only male participants in this study. Third, the high level of engagement within liberal arts institutional culture is substantially different from large universities and, therefore, the results may be subject to limitations regarding the possible effects of engagement level. Finally, the study did not include subjective perceptions of participants about their performance or the method used.

Conclusion

We investigated the effectiveness of the flipped classroom approach for improving performance on a high stakes test. We contribute to the literature by applying the pedagogy in growth theory and implementing it in an intermediate class with small, highly interactive sections. While keeping instructional capital constant, our results suggest that flipped pedagogy, in general, did not improve outcomes. There is a weak exception for high achieving students with high SAT scores. The most important determinant of good performance was the quality of the students' high school. This reflects available resources and, thus, the socio-economic environment of the students. Our results align with those of Craft and Linask (2019) and Green (2014), and point to the need for further investigation of the role of pre-class material as in Webb et al. (2021). Additionally, the results suggest that students from non-marginalized groups from resource-rich high schools are more likely to benefit from the approach, as suggested by Ficano (2019) and Setren et al. (2020). This underscores the need for careful preparation of flipped pedagogy, especially with regard to setting up a level playing field for all students. The approach should not be undertaken lightly without careful advance thinking about its implementation. Similarly, there is an acute need for appropriate preparation of students, especially if some of them come from less privileged backgrounds.

Appendix 1: High school rating

See Fig. 4.

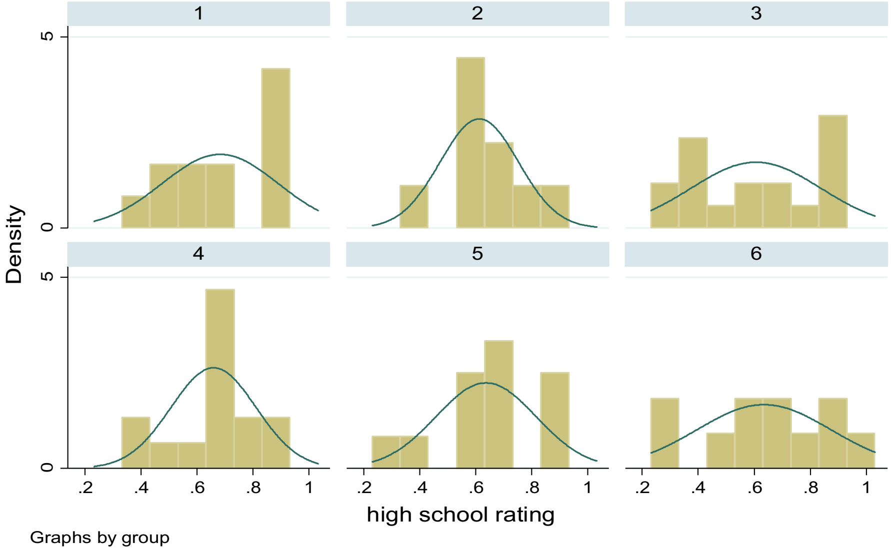

Fig. 4 High school rating by individual sections (treatment groups are sections "We are grateful to D. Dalenberg for assistance with the initial raw data and explanation of high school rating." and 6)

Appendix 2: The Instrument

ECO 292: Intermediate Macroeconomics
Date

Exam

This is a timed exam. You have 30 minutes to complete it. Please use your time wisely. Good luck!

1. For the following information $Y=3K^{1/5}L^{4/5}$, $s = 1/10$, $d = 0.13$, $n = 0.03$, $g = 0.03$ find
    a. the initial steady state and

- b. highest possible sustainable consumption per effective worker
- c. imagine now that saving rate increases to 1/5
    - I. Draw a graph to illustrate this change
    - II. Calculate the highest sustainable consumption now.
    - III. By how many % did it (the highest sustainable consumption) change?
    - IV. How much is growth of output per worker in the new steady state?

2. Consider an economy where real GDP on average grows at 3%. Depreciation rate in this economy is 5%, value of capital is 3 times annual GDP, and capital share of income is 25%. Assume Cobb-Douglass production function and economy in a steady state.
    - a. Find initial saving rate for this economy
    - b. Determine if the economy is initially at the level of capital per effective worker that provides highest sustainable consumption.
    - c. Find capital-output ratio for the Golden Rule steady state
3. There are two blanks in the table below. Show how the missing entries are calculated. This reproduces table 8-2 on page 228 of your textbook but changes the production function to $Y = (1/2)K^{1/3}L^{2/3}$.

Exogenous variables						
y	0.5*k^1/3					
Initial k	4					
s	0.3					
d	0.1					
Endogenous variables						
Output/income per worker	y					
Capital stock	k					
Investment per worker	i					
Depreciation	(d+n+g)k					
Consumption per worker	c					
Year	k	y	c	i	$(d+n+g)k$	*deltak*
1	4.000	0.794	0.556	0.238	0.400	−0.162
2	3.838	0.783	0.548	0.235	0.384	−0.149
3	3.689	0.773	–	0.232	0.369	−0.137
4	3.552	0.763	0.534	0.229	0.355	−0.126
5	3.426	0.754	0.528	0.226	0.343	−0.116
6	3.309	0.745	0.522	0.224	0.331	−0.107
7	3.202	0.737	0.516	–	0.320	−0.099
8	3.103	0.729	0.510	0.219	0.310	−0.091
9	3.011	0.722	0.505	0.217	0.301	−0.085

4. The book presents a case study: “The Miracle of Japanese and German Growth.” Use the Solow canonical graph to illustrate these two cases.
5. Super concisely explain the gist of the Kremerian model discussed in your textbook.

References

Acemoglu, D. 2013. Economic growth and development in the undegraduate cuirriculum. *The Journal of Economic Education* 44(2): 169–177.

Akcayir, G., and M. Akcayir. 2018. The flipped classroom: A review of its advantages and challenges. *Computers and Education* 126: 334–345. https://doi.org/10.1016/j.compedu.2018.07.021.

Balaban, R.A., D.B. Gilleskie, and U. Tran. 2016. A quantitative evaluation of the flipped classroom in a large lecture principles of economics course. *The Journal of Economic Education* 47(4): 269–287. https://doi.org/10.1080/00220485.2016.1213679.

Ballard, C.L., and M.F. Johnson. 2004. Basic math skills and performance in an introductory economics class. *Journal of Economic Education* 35(1): 2–23. https://doi.org/10.3200/JECE.35.13-23.

Barreto, H. 2015. Why excel? *The Journal of Economic Education* 46(3): 300–309. https://doi.org/10.1080/00220485.2015.1029177.

Barreto, H. 2016. Teaching Macroeconomics with Excel. Accessed 09 15, 2016. https://www.depauw.edu/learn/macroexcel/screencasts/index.htm.

Benge, M., and G. Wells. 2002. Growth and the current account in a small open economy. *The Journal of Economic Education* 33(2): 152–165. https://doi.org/10.1080/00220480209596464.

Bolt, Juta, and J.L. vanZanden. 2020. Madison Project Database 2020. University of Groningen Groningen Growth and Development Centre. Accessed 12 27, 2022. https://www.rug.nl/ggdc/historicaldevelopment/maddison/releases/maddison-project-database-2020?lang=en.

Bosshardt, W., and N. Manage. 2011. Does calculus help in principle of economics courses? Estimates using matching estimators. *The American Economist* 56(1): 29–37. https://doi.org/10.1177/056943451105600105.

Boyle, A., and W. Goffe. 2018. Beyond the flipped class: the impact of research-based teaching methods in a macroeconomics principle class. *AEA Papers and Proceedings* 108: 297–301. https://doi.org/10.1257/pandp.20181052.

Calimeris, L., and K.M. Sauer. 2015. Flipping out about the flip: All hype or is there hope? *International Review of Economics Education* 20: 13–28. https://doi.org/10.1016/j.iree.2015.08.001.

Card, D., and J. Rothstein. 2007. Racial segregation and the black-white test score gap. *Journal of Public Economics* 91(11–12): 2158–2184. https://doi.org/10.1016/j.jpubeco.2007.03.006.

Caviglia-Harris, Jill. 2016. Flipping the undergraduate economics classroom: Using online videos to enhance teaching and learning. *Southern Economic Journal* 83(1): 321–331. https://doi.org/10.1002/soej.12128.

Chen, J., and T.-F. Lin. 2012. Do supplemental online recorded lectures help students learn microeconomics. *International Review of Economics Education* 11(1): 6–15. https://doi.org/10.1016/S1477-3880(15)30021-9.

Cosgrove, S.B., and N.H. Olitsky. 2020. Research-based instructional strategies in a flipped principles of microeconomics classroom. *International Review of Economic Education* 33: 100175. https://doi.org/10.1016/j.iree.2020.100175.

Craft, E., and M. Linask. 2019. Learning effects of the flipped clasroom in a principles of microeconomics course. The Journal of Economic Education, 1-18.

Elmslie, B.T., and E. Tebaldi. 2010. Teaching economic growth theory with data. *The Journal of Economic Education* 41(2): 110–124. https://doi.org/10.1080/00220481003617244.

Estes, M.D., R. Ingram, and J.C. Liu. 2014. A review of flipped classroom research, practice, and technologies. International HETL Review 4 (7). https://www.hetl.org/a-review-of-flipped-classroom-research-practice-and-technologies/.

Ficano, C.K.C. 2019. Indentifying differential benefits from a flipped-group pedagogy in introductory microeconomics. *International Review of Economics Education*. https://doi.org/10.1016/j.iree.2018.07.002.

Flores, N.J., and Scott Savage. 2007. Student demand for streaming lecture video: Empirical evidence from undergraduate economics classes. *International Review of Economics Education* 6(2): 57–78. https://doi.org/10.1016/S1477-3880(15)30103-1.

Freeman, S., D. Haak, and M.P. Wendroth. 2011. Increased course structure improves performance in introductory biology. *Life Science Education* 10(2): 175–186.

Green, A. 2014. The case for the traditional classroom. *International Review of Economics Education* 16: 87–99. https://doi.org/10.1016/j.iree.2014.04.002.

Gulley, O.D., and A.L. Jackson. 2016. A case study on using instructor-recorded videos in an upper level economics course. *International Review of Economics Education* 23: 28–33. https://doi.org/10.1016/j.iree.2016.06.004.

Haak, D.C., J. HilleRisLambers, E. Pitre, and S. Freeman. 2011. Increases structure and active learning reduce the achievemetn gap in introductory biology. *Science* 332(6034): 1213–1216. https://doi.org/10.1126/science.1204820.

Hettler, P.L. 2015. Active Learning in economics: Increasing student engagement, excitement and success. *International Advances in Economic Research* 21: 357–360. https://doi.org/10.1007/s11294-015-9548-6.

Lage, M. J., G. J. Platt, and M. Treglia. 2000. Inverting the classroom: A gateway to creating an includsive learning environment. Journal of Economic Education 30-43.

Lombardini, C., M. Lakkala, and H. Muukkonen. 2018. The impact of the flipped classroom in a principles of microeconomics course: Evidence from a quasi-experiment with two flipped classroom designs. *International Review of Economics Education* 29: 14–28. https://doi.org/10.1016/j.iree.2018.01.003.

Mankiw, N.G. 2019. Intermediate micro and macro: Six guidelines for teaching intermediate macroeconomics. *The Journal of Economic Education* 50(3): 258–260. https://doi.org/10.1080/00220485.2019.1618768.

Mankiw, N.G. 2022. *Macroeconomics*. New York: Worth Publishers.

Mendez-Carbajo, D., and L.C. Malakar. 2020. Flipping the classrooom with Econlowdown.org. *The Journal of Economic Education* 51(1): 95–102. https://doi.org/10.1080/00220485.2019.1687376.

Mu, H., and D. Paparas. 2016. Ready for the flipped classroom? Preliminary experiences of the new approach in teaching economics to non-major students. *Applied Economics and Finance* 3(2): 45–53. https://doi.org/10.11114/aef.v3i2.1288.

Purba, S.E.E., K. Kristiani, K.B. Sangka, and O.K. Hussain. 2021. The flipped clasroom: An overview of tis impact on economics learning. *International Journal of Pedagogy and Teacher Education* 5(1): 26–34. https://doi.org/10.20961/ijpte.v5i1.49750.

Roach, T. 2014. Student perceptions toward flipped learning: New methods to increase interaction and active learning in economics. *International Review of Economic Education* 17: 74–84. https://doi.org/10.1016/j.iree.2014.08.003.

Setren, E., K. Greenberg, O. Moore, and M. Yanukovich. 2020. Effects of the flipped clasroom: Evidence from a randomized trial. *Education Finance and Policy*. https://doi.org/10.1162/edfp_a_00314.

Singh, N. 2020. A little flip goes a long way: The impact of a flipped clasroom design on student performance and engagemetn in a first-year undergraduate economics clasroom. *Education Science* 10(11): 1–15. https://doi.org/10.3390/educsci10110319.

Stein, S.H. 2007. A beginner's guide to the solow model. *The Journal of Economic Education* 38(2): 187–193.

Strelan, P., A. Osborn, and E. Palmer. 2020. The flipped classroom: A meta-analysis fo effects on student performance across disciplines and education levels. *Educational Research Review*. https://doi.org/10.1016/j.edirev.2020.100314.

Swoboda, A., and L. Feiler. 2016. Measuring the eEffect of blended learning: Evidence from a selective liberal arts college. *American Economic Review* 106(5): 368–382. https://doi.org/10.1257/aer.p20161055.

Vazquez, J.J., and E.P. Chiang. 2015. Flipping out! A case study on how to flip the principles of economics classroom. *International Advances in Economic Research* 21: 379–390. https://doi.org/10.1007/s11294-015-9549-5.

Webb, R., D. Watson, C. Shepherd, and S. Cook. 2021. Flipping the classroom: Is it the type of flipping that adds value? *Studies in Higher Education* 46(8): 1649–1663. https://doi.org/10.1080/03075079.2019.1698535.

Wozny, N., C. Balser, and D. Ives. 2018. Evaluating the flipped clasroom: A randomized controlled trial. The Journal of Economic Education, 115-129.

Yamarik, S. 2019. Flipping the classroom and student learning outcomes: Evidence from an international economics course. *International Review of Economics Education*. https://doi.org/10.1016/j.iree.2019.100163.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.