RESEARCH ARTICLE

Open Access

Investigating effects of teachers in flipped classroom: a randomized controlled trial study of classroom level heterogeneity

Julie Buhl-Wiggers[1], Lisbeth la Cour[1], Mette Suder Franck[1] and Annemette Kjærgaard[2*]

*Correspondence:
amk.msc@cbs.dk

[1] Department of Economics, Copenhagen Business School, Frederiksberg, Denmark
[2] Department of Management, Society and Communication, Copenhagen Business School, Dalgas Have 15, 2000 Frederiksberg, Denmark

Abstract

The increased popularity of flipped classroom in higher education warrants more thorough investigation of the pedagogical format's effects on student learning. This paper utilizes two iterations of a randomized field experiment to study the effects of flipped classroom on student learning specifically focusing on heterogeneous treatment effects across the important classroom-level factor of teachers. The empirical setting is an undergraduate macroeconomics course with 933 students and 11 teachers. Our findings show a positive yet insignificant average effect of flipped classroom on both pass rate and final exam grades. We further find substantial shifts in the ranking of the participating teachers' effectiveness when comparing traditional and flipped classroom conditions, which suggests that the most successful teacher in a traditional teaching environment is not necessarily the most successful teacher in a flipped classroom environment.

Keywords: Flipped classroom, Teacher heterogeneity, Randomized controlled trial, Academic achievement, Economics education

Introduction

Knowledge about teaching and learning in flipped classroom has grown substantially in the last decades, as educational scholars have mirrored the rising interest in the pedagogical format displayed by teachers with an interest in supporting active learning (Lai et al., 2021; Sun et al., 2018). Arguably, a significant reason for the increased interest in flipped classroom is the recent technological development that has enabled a shift of content traditionally delivered in-class to an online out-of-class setting (Lai, 2021). This has freed up in-class time for more student-centered activities (Bergmann & Sams, 2012; McLaughlin et al., 2014; O'Flaherty & Phillips, 2015), as for example increased student interaction with teachers (van Alten et al., 2019).

The increased popularity of flipped classroom is reflected in the number of empirical studies aimed at assessing the potential of the format to benefit student learning outcomes, such as test scores or exam grades. In general, meta-analyses of this body of literature suggest a potential for flipped classroom to benefit student learning outcomes (O'Flaherty & Phillips, 2015; Strelan et al., 2020). At the same time there is a

non-negligible number of studies that do not find conclusive evidence that flipped classroom is superior to traditional teaching formats (see for example Chen Hsieh et al., 2017; Love et al., 2014; Nielsen et al., 2018; Mingorance et al., 2019). A potential explanation of the ambiguous results of these studies is the presence of underlying heterogeneity. A number of papers on flipped classroom consider this possibility by assessing whether the effect of the format varies according to student-level characteristics (Ficano, 2019; Nouri, 2016; Ryan & Reid, 2016), but only limited attention has been dedicated towards investigating heterogeneity related to teachers.

The change in in-class activities within flipped classroom suggests that factors at the classroom-level, such as teachers, may have a different influence on student performance compared to traditional classrooms (Brewer & Movahedazarhouligh, 2018; Kim et al., 2014). Indeed, research has indicated that successful teaching using flipped classroom involves substantially different skill sets than those demanded in traditional classrooms and that student-centered learning requires teachers to reconsider their role and way of teaching when they engage in flipped classroom (see for example Akçayır & Akçayır, 2018; Lai & Hwang, 2016; Sun et al., 2018).

Despite acknowledgement of the changing roles of teachers in flipped classroom, the influence of teachers has to our knowledge not been the main focus of quantitative investigations. Accordingly, we explicitly explore how the effect of flipped classroom varies across teachers. The research questions guiding the study are:

1. Does the effect of flipped classroom vary between teachers?
2. Do teachers maintain their effectiveness rank when switching to flipped classroom?

To pursue these questions, we apply a quantitative approach to study the effect of a pedagogical intervention inspired by flipped classroom, designed as a randomized controlled trial (RCT) and implemented in the tutorial classes of a macroeconomics course at a large Scandinavian business school. The intervention was first introduced in 2018 and repeated in 2019. Since only a few teachers taught the course both years, we pool the two iterations of the RCT to increase the number teachers considered. This leaves us with an analytical data set of 11 teachers and 933 students.

The contribution of this study is to go beyond the sheer comparison of student learning outcomes in traditional vs flipped classroom and instead explore how teachers influence the effect of flipped classroom on student performance. The findings show variability in the success of flipped classroom in terms of increasing student performance across teachers. We observe several cases of relative rank reversals in teacher effectiveness between the two formats. This provides some quantitative empirical evidence corroborating the notion from previous qualitative research (e.g. Akçayır & Akçayır, 2018) that teacher skills required to ensure good student performance in flipped classroom are distinct from those of traditional classrooms.

The paper proceeds as follows. First, we provide an overview of the literature on flipped classroom in general and heterogeneous effects in particular. Then we describe the design of the study, the data collected and the empirical strategy. The results are presented next, followed by a discussion, before we conclude and outline some limitations of the study with suggestions for further research.

Literature review

Studies concerned with assessing the potential of flipped classroom to increase student outcomes in higher education have reported somewhat mixed results. In a recent meta-analysis, Strelan et al. (2020) find an average effect size for student performance of 0.48 SD for higher education, yet this varies significantly with discipline; for example, Lo and Hew (2019) found positive effects in a meta-analysis of engineering education, while no significant effect was found in a systematic review of medical education (Chen et al., 2017). Similarly, the strand of research in economics education to which the present study belongs, report marked differences in their estimates of the average effect of flipped classroom. While findings by Calimeris and Sauer (2015) show that flipped classroom increases students' average performance on the final exam by 0.64 standard deviations, other studies find no statistically significant effect on the final exam (Setren et al., 2021), and Wozny et al. (2018) only find a positive effect on the final exam for high-achieving students.

While the existing literature suggests some explanations for differences in the effect of flipped classroom, the scope for investigating such heterogeneity has predominantly been limited to the characteristics of students (see for exampleNouri, 2016; Ryan & Reid, 2016; Ficano, 2019) and little attention has been put on teachers as a source of heterogeneity in the effect of flipped classroom on student performance. At the general level, teachers are widely acknowledged among educational economists as central for students' academic success (for example noted in Hanushek & Rivkin, 2006). Teachers' effects on student achievement have most often been studied at lower levels of education and within this context, studies assessing the effect of observable teacher characteristics, such as education and certification, report mixed results (Carrell & West, 2010). However, several studies computing a measure of total teacher effectiveness capturing both observed and unobserved factors find that teacher quality has notable effects on students' test scores (Kane & Staiger, 2008; Rivkin et al., 2005; Rockoff, 2004).

While teachers are frequently mentioned as being important for student learning in discussions on flipped classroom more broadly, they are rarely the primary focus of studies (see Appendix Table 5 for an overview of the articles mentioned here). One example where teachers do appear as part of a study's explicit objective of identifying factors conducive for successful implementation of flipped classroom, is in the qualitative study by Kim et al., (2014). The study combines a range of empirical data such as student surveys, interviews, and instructor reflections to outline what aspects of flipped classroom are especially beneficial for teaching and learning. Based on their analyses, the authors formulate design principles including a strong emphasis on the teacher's role as facilitator to ensure student engagement. The importance of "Teacher Presence" is evident in students' wish for well-structured and clearly defined guidance for concrete assignments but also for supporting student interactions and facilitating collaborative learning (Kim et al., 2014).

When teachers' roles are addressed in the flipped classroom literature, it is often regarding increased workload due to changing the format of courses (Karabulut-Ilgu et al., 2018). Another frequent mentioning is how flipped classroom is more closely

related to teachers' pedagogical impact by for example arguing that teachers' role in flipped classroom is distinct from traditional classrooms (Akçayır & Akçayır, 2018; DeLozier & Rhodes, 2017). Other studies note that specific teaching beliefs are a prerequisite for successful flipped teaching (Hwang et al., 2015), or that teachers need to provide individualized student instruction and scaffolding during in-class activities (Ghadiri, 2014). Similarly, some authors argue that shifting to student-centered learning in flipped classroom changes the role of teachers towards facilitation of learning rather than transmission of knowledge and moves part of the responsibility for learning from teachers to students (Zou et al., 2020). This suggests that teachers' implementation of the format is pivotal for benefitting student learning outcomes (DeLozier & Rhodes, 2017). Nevertheless, despite such seeming consensus of acknowledging teachers' importance in flipped classroom, the literature is surprisingly void of quantitative empirical studies investigating if the effect of the pedagogical format varies across teachers. In the following sections, we therefore zoom in on the classroom and examine teacher heterogeneity in a pedagogical format inspired by flipped classroom. We begin by outlining the details of our setting and RCT.

Setting and experimental design

The intervention investigated in the present study, took place in a second semester introductory macroeconomics course at the largest study program in a Scandinavian business school targeting the tutorial classes of two consecutive cohorts: 14 classes in 2018 and 15 classes in 2019 with approximately 45 students in each. The tutorials were scheduled to 90 min a week and participation was voluntary as is national standard regulation for university education. In the traditional framework, students were expected to work with assigned exercises before attending the tutorial classes with the intention of freeing up space in-class for students to ask clarifying questions. However, students often came to class un(der)prepared making the tutorials highly teacher-centered and more like "mini-lectures". Consequently, it was decided to make the activities of these classes more student-centered, and this change of format is the focus of the RCT.

Intervention design

The setup was motivated by the flipped classroom idea of increasing in-class activity in the tutorial classes while lectures proceeded as usual. In this respect, the intervention deviated from a standard flipped classroom setting where lectures are often provided online before in-class tutorials. More specifically, the overall aim of the intervention was to rely on the flipped classroom philosophy of freeing up time for more student-centered learning. Half of the tutorial classes were changed to a new, more active format (treatment), while the other half formed a business-as-usual control group (control). The intervention was introduced to students through an information e-mail and in-class presentation in the week prior to the beginning of the semester. Students had the opportunity to opt out of the research by withdrawing consent to the use of their data and the research project was approved by the institutional ethical review board.

The treatment group engaged in collaborative group work on a weekly assigned problem set. Instructors facilitated the group work and supported students during problem

solving exercises. To ensure correction of misconceptions, the treatment group had access to video solutions to the assigned problem set after class. In the control group, students were supposed to engage with solving the problem set out-of-class, while the teacher explained the solutions in-class. These students did not have access to the video-solutions. Finally, and of particular importance for the teacher focus of this paper, the teachers were carefully prepared to teach the new format by participating in workshops before the start of the semester. Members of the business school's pedagogical unit were engaged in these preparatory workshops. Table 1 provides an overview of the intervention.

Randomization procedure

When students at the business school are enrolled in a specific study program, they are stratified by gender and nationality and randomly assigned to tutorial classes. One exception is that the older students are placed in the same tutorial classes. In both intervention years, we made use of randomization to measure the impact of the intervention, however, the level of randomization differed between the two years. In 2018 we randomized at the student-level, thus randomly placing each individual student in either a treatment or a control group and subsequently divide the treatment and control group into 7 tutorial classes. In the 2019 iteration, it was decided not to break up the pre-assigned tutorial classes and therefore randomization was at the tutorial class-level instead.

In both years, students in the treatment group were assigned to tutorial classes but not to specific study groups within the classroom. This meant that students self-selected into study groups without any interference by the teacher, unless one or more students did not have any peers to collaborate with in which case the teacher would allocate students to study groups.

To ensure that our results were not affected by potential differences in teachers' competence, we stratified the treatment assignment by teacher in both years, so that each teacher taught both a treatment and a control class. To address potential time-of-day effects, all classes were scheduled for the same day. Because each teacher taught two

Table 1 Overview of intervention

	Control	Treatment
Lectures		
Before class	Textbook readings	Textbook readings
In class	Large class lecture: 90 min per week	Large class lecture: 90 min per week
Tutorial classes		
Before class	Prepare problem set*	
In class	Students listen and ask questions, while teacher explains problem set. Teacher gives a "mini lecture"	Students participate in group work on problem set. Teacher acts as a facilitator
After class		Students get access to video solutions of problem sets
Exam		
	4 h closed book exam	4 h closed book exam

*Although students in the control classes were expected to prepare before class, many students showed up to the tutorial class without preparing

classes, not all classes could be placed in the same time slot. Therefore, they were placed back-to-back and time slots switched halfway through the course. A research assistant monitored access to the classrooms to ensure that only students assigned to the treatment classes gained access. Similarly, access to online materials was limited to the treatment group through the learning management system.

Data

Student performance in the macroeconomics course was assessed only once at a final closed-book exam. Grading was based on an absolute grading system, blinded, and performed by an internal teacher, who randomly received a subset of exams from all of the different tutorial classes. To assess the effect of the intervention, two main outcomes are considered: (1) The grade from the final exam, which was standardized by the mean and standard deviation of the control group in each year, and (2) a binary pass/fail measure, where fail include both failing grades and no-shows. The business school's own administrative data provide information on the two outcomes as well as on a number of student-level variables that are included as controls in the analyses; two separate ability measures, age, gender, enrollment year, and whether they participated in the re-take exam in the fall course in microeconomics. Age is measured in years and the three latter variables are defined as dummy variables. We include information on the students' potential participation in the re-take of the microeconomics exam because the timing of this exam coincided with the beginning of the macroeconomics course. Therefore, students who participated in this re-take exam might have had a more challenging start to the macroeconomics course than those who did not. The two ability measures are high school GPA and an ECTS weighted GPA from the fall semester immediately before the intervention took place. Both measures are included as controls because we expect them to capture distinct abilities. High school GPA reflects academic capability in a range of diverse subjects and for this reason also provides an indication of motivation and diligence. On the other hand, the GPA from the fall semester constitutes a quantitative measure of the students' performances in economics specific courses, as well as their adaption to the teaching and exam formats at the university.

From the full sample of 1215 students and 13 teachers, our analytical sample was obtained in the following way. First, the sample was restricted to only include students who participated in at least one exam during the first semester and did not drop out during the second semester while the intervention took place. Second, we identified and excluded students in the control group who gained access to the online video solutions removing them from our sample. We did this to address potential spillover effects. Two teachers were dropped from our analytical sample as the one only taught one class which did not allow us to control for teacher fixed effects. The other taught the two classes in 2019 with older students that were exempted from the randomization. Finally, we only included students in our estimation samples for whom we have information on their high school GPA. These restrictions resulted in a dataset comprising 11 teachers and 933 students (509 in 2018 and 424 in 2019) of which 763 (415 in 2018 and 348 in 2019)

participated in the final exam in macroeconomics. Appendix Fig. 5 summarizes the process of data cleaning.

Measuring teacher effects

Since each teacher in the estimation sample taught at least one treatment and one control tutorial class, we can control for a teacher's average "teacher effect" by including teacher fixed effects in our regressions. This reduces the risk of confusing treatment and teacher effects. In practice, this is achieved by including a dummy for all but one teacher. The design also allows to estimate the treatment effect separately for each teacher and thereby shed light on the heterogeneity of treatment effects across teachers.

Balance and descriptive statistics

Table 2 presents balance on pre-treatment observable characteristics between the students in the treatment and control group for both the full analytical sample and the sample with grades on the final exam. The two samples differ in size, because the analyses with the binary pass variable as the outcome also include students who did not show up for the exam and therefore did not receive a grade. Table 2 shows no issues of imbalance for the variables.

Overall, there are no unexpected differences in the descriptive statistics between the two samples (Table 3). GPA from prior semester is lower and the share participating in the microeconomics retake exam higher for the full analytical sample. This is no surprise, as students who did not participate in the macroeconomics exam are arguably also more likely not to have participated in previous exams than the students who did. The sample with grades on the final exam shows that students on average received a grade of 6.11 in macroeconomics, which is very close to the sample average of the weighted prior semester GPA of 6.17. Table 3 further shows that the mean age is 21.2 years and that the study program has most male students.

Table 2 Balance of pre-treatment covariates between treatment and control group

	Full sample (N = 933)			Sample w. exam grade (N = 763)		
	Control (N = 457)	Treatment (N = 476)	p-value	Control (N = 367)	Treatment (N = 396)	p-value
Female	0.337	0.319	0.543	0.357	0.311	0.154
	(0.022)	(0.021)		(0.025)	(0.023)	
Age	21.306	21.265	0.579	21.191	21.205	0.929
	(0.090)	(0.062)		(0.088)	(0.065)	
High School GPA	8.987	8.996	0.908	9.017	9.071	0.551
	(0.062)	(0.058)		(0.070)	(0.061)	
GPA Fall	5.698	5.801	0.556	6.123	6.219	0.605
	(0.126)	(0.126)		(0.131)	(0.1319	
Retake microeconomics exam	0.302	0.326	0.462	0.232	0.245	0.715
	(0.021)	(0.022)		(0.0229	(0.022)	

Displays mean values with standard errors in parentheses. P-values reflect the t-tests of equality of means across treatment and control. The full sample includes all students remaining after the cleaning process, while the sample with exam grade excludes students who did not participate in the final exam of the macroeconomics course

Table 3 Descriptive statistics

	Full sample (N = 933)		Sample w. exam grade (N = 763)	
	Mean	SD	Mean	SD
Outcome				
Macroeconomics grade			6.11	3.85
Pass rate	0.72	0.45	0.89	0.32
Controls				
Age	21.29	1.65	21.20	1.49
Female	0.33	0.47	0.33	0.47
High School GPA	8.99	1.29	9.05	1.27
GPA Fall	5.75	2.72	6.17	2.56
Retake microeconomics exam	0.31	0.46	0.24	0.43

The full sample includes all students remaining after the cleaning process, while the sample with exam grade excludes students who did not participate in the final exam of the macroeconomics course

Empirical strategy

To assess the overall effect of the intervention on student outcomes, we begin our analysis by looking at the average treatment effect, which we estimate by a pooled OLS regression:

$$Y_{icky} = \beta_0 + \beta_1 T_i + \varepsilon_{icky} \tag{1}$$

where i denotes the individual student, c the student's classroom, k teacher, and y the year of participation in the macroeconomics course. Y_{icky} is either the pass rate, or the standardized grade from the final exam. T_i is a dummy variable taking the value of one if the student was enrolled in a tutorial affected by the intervention and zero otherwise. To increase precision we estimate an augmented version of Eq. (1) that includes a number of covariates; a year dummy, D19, which takes the value of one if the student was enrolled in the course in 2019 and zero if enrollment was in 2018, which allows for differences in the effect of the intervention, age, gender, high school GPA, previous semester GPA and teacher fixed effects.

In all our analyses, we consider it likely that there might be intra-class correlation of the outcomes within tutorial groups or of students taught by the same teacher, as they are exposed to the same learning environment. Consequently, the regression tables in this paper report p-values based on a wild cluster bootstrap (WCB) procedure for inference, which is the common approach to addressing intra-class correlation in empirical settings with few clusters (Colin Cameron & Miller, 2015). We cluster at the tutorial class level.

After the analysis of the average treatment effect, we turn towards answering the two research questions and examine if and how the effect of the intervention varies among the 11 teachers. To answer the first research question, we estimate a model including all covariates from the full estimation model of the average treatment effect and additionally include an interaction term between treatment and each teacher dummy:

$$Y_{icky} = \beta_0 + \beta_1 T_i + \beta_2 D19_{iy} + X'_{i\rho} + \delta_1 Teacher_1 + \ldots$$
$$+ \delta_{10} Teacher_{10} + \gamma_1 T_i * Teacher_1 + \ldots + \gamma_{10} T_i * Teacher_{10} + \varepsilon_{icky} \quad (2)$$

Here our main interest lies in assessing the coefficients, γ_k, on the interactions between treatment status and each teacher. The coefficient estimates of these interactions inform us about whether the average outcome of students in the teacher's treatment class(es) is different from that of the teacher's control class(es).

Because estimation of Eq. (2) provides us with estimates of the difference in outcomes between a teacher's treatment and control class(es), it does not allow us to assess a teacher's effect on average student outcomes in each setting. Therefore, to gain further insights on the relationship between the intervention and teacher effectiveness we follow a procedure suggested by McCaffrey et al. (2012). This methodological approach helps us answer the second research question by offering a way to compute separate mean corrected estimates of the average grades and pass rates of the students in the control and treatment classes for each of the teachers. More specifically, the teacher effectiveness in each classroom setting is computed as the mean outcome of a teacher's students (after correcting for the effect of other regressors) minus the overall corrected mean for all students. We then use these measures as the basis for computing the teachers' effectiveness ranks separately for the flipped and traditional classrooms.

In the analyses of teacher heterogeneity, we cannot rely on WCB standard errors at the class level due to an insufficient number of clusters to compute cluster-robust WCB for the teacher fixed effects. Consequently, we instead rely on heteroscedasticity robust standard errors.

Table 4 Average treatment effects

	Pass rate			Exam grade		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.035	0.036	0.037	0.100	0.065	0.059
	[0.322]	[0.214]	[0.270]	[0.192]	[0.346]	[0.330]
Experiment year		0.040	0.023		− 0.020	− 0.173**
		[0.244]	[0.568]		[0.752]	[0.026]
Fall GPA		0.054***	0.054***		0.250***	0.252***
		[0.000]	[0.000]		[0.000]	[0.000]
High school GPA		− 0.000	0.001		0.076***	0.072***
		[0.968]	[1.000]		[0.002]	[0.002]
Female		0.002	0.001		− 0.145**	− 0.152**
		[0.946]	[0.996]		[0.018]	[0.014]
Age		− 0.008	− 0.009		0.013	0.015
		[0.398]	[0.314]		[0.404]	[0.292]
Retake microeconomics exam		− 0.239***	− 0.238***		− 0.053	− 0.045
		[0.000]	[0.000]		[0.462]	[0.486]
Observations	933	933	933	763	763	763
R-squared	0.002	0.274	0.282	0.003	0.502	0.516
Teacher fixed effects	No	No	Yes	No	No	Yes

p-values in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1. Uses wild cluster bootstrap standard errors. Clustered at the class level with B = 1000

Results

Average treatment effects

Table 4 presents the results from estimating Eq. (1) for the two outcomes of interest. Column (1) and (4) show the raw average treatment effect for the pass rate and exam grade, respectively. Column (2) and (5) add controls for increased precision, while Column (3) and (6) additionally includes teacher fixed effects.

Although the coefficient estimates on the treatment dummy suggest a positive treatment, the estimated effect of the flipped classroom intervention is insignificant across all model specifications. This is largely consistent with previous studies of the average treatment effect of flipped classroom in teaching and learning within the field of economics (as e.g. reported by Setren et al., 2020 and Wozny et al., 2018).

For experiment year, age, high school GPA, and gender we see no significance for the pass outcome. However, for the exam grade the coefficient of the experiment year is significantly negative, when we control for teacher fixed effects, as is the coefficients on gender regardless of inclusion of these fixed effects. Unsurprisingly, in all regressions the student's GPA from the fall semester is estimated to be a positive and significant predictor of performance in the macroeconomics exam. Moreover, for the pass rate, our results indicates that students who participated in the retake exam in microeconomics are significantly less likely to pass the macroeconomics exam. For the exam grade, we see no significance for this variable. Finally, for the exam grade, we also find significant
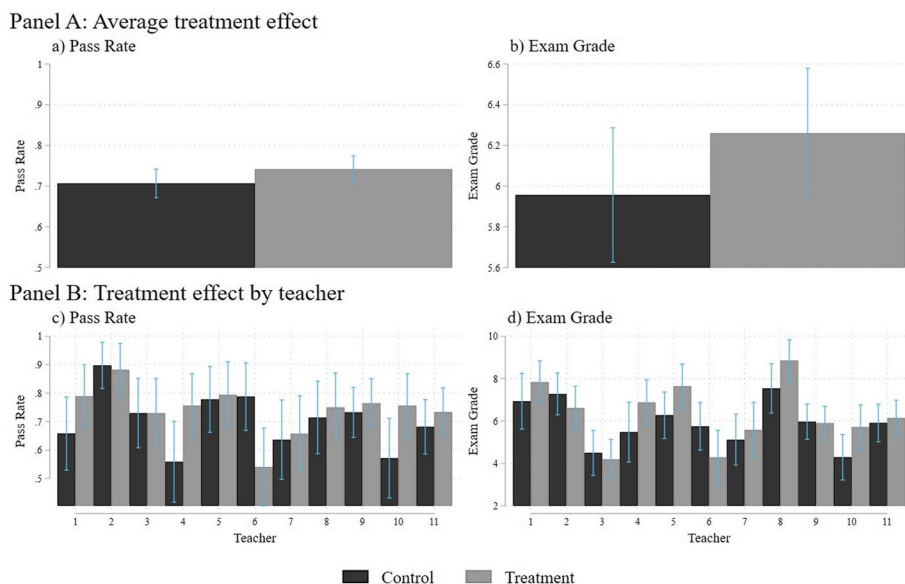

Fig. 1 Average treatment effects and teacher heterogeneity. Bars indicate 90% confidence intervals based on heteroscedasticity robust standard errors. Contrary to Table 4, these graphs show the level of the exam grades rather than the standardized grades used in our regression models. Consequently, the difference between the average values of the treatment and control groups do not match the coefficient estimate on the treatment dummy reported in Column (4) of Table 4. Display raw differences without inclusion of controls. For the pass rate outcome the test of equality of means across treatment status for each teacher suggests a significant difference in the raw means for three teachers, namely Teacher 4 (p = 0.077), Teacher 6 (p = 0.028), and Teacher 10 (p = 0.094). For the exam grade outcome none of the raw differences are significantly different from zero at a 10 percent significance level

and positive effects for high school GPA though the magnitude of this effect is notably smaller than for the fall GPA. This suggests that a student's performance in higher education economics-specific courses is a better predictor of their exam grade in the macroeconomics exam than the broader measure of previous academic achievements and diligence that we attempt to capture by the high school GPA.

Because teachers, as mentioned in the literature review, are widely acknowledged as being central to students' educational outcomes, variation in the effect of the flipped classroom intervention across teachers might explain why we do not find a significant average treatment effect. Figure 1 plots the average pass rate and exam grades for students by treatment status (subplot a) and (b) in Panel A and by both treatment status and teacher (subplot c) and (d) in Panel B. This figure offers some explorative insights on whether our finding of no significant effect of the intervention could be explained by classroom-level heterogeneity due to teachers. Panel A shows the modest differences in the raw treatment effects, while Panel B indicates marked differences in students' average performances in their macroeconomics exam between students taught in traditional classrooms and flipped classroom, when making within-teacher comparisons. For the pass rate outcome displayed in subplot (c), the within-teacher difference is most clearly pronounced for Teacher 1, 4, and 10, where the average pass rate of students in the control group is considerably lower than in the treatment group. However, for Teacher 6, the average pass rate of students in the control group greatly exceeds those of the students in the flipped classroom setting. Similarly, the within-teacher comparisons of the average exam grade displayed in subplot (d) also suggest some cases of notable differences, namely for Teacher 4, 5, 8, and 10.

Overall, Fig. 1 provides some informal indications that teacher heterogeneity might constitute a source of heterogeneity in the effect of the flipped classroom intervention. This motivates the formal exploration of teacher heterogeneity, which we turn to next.
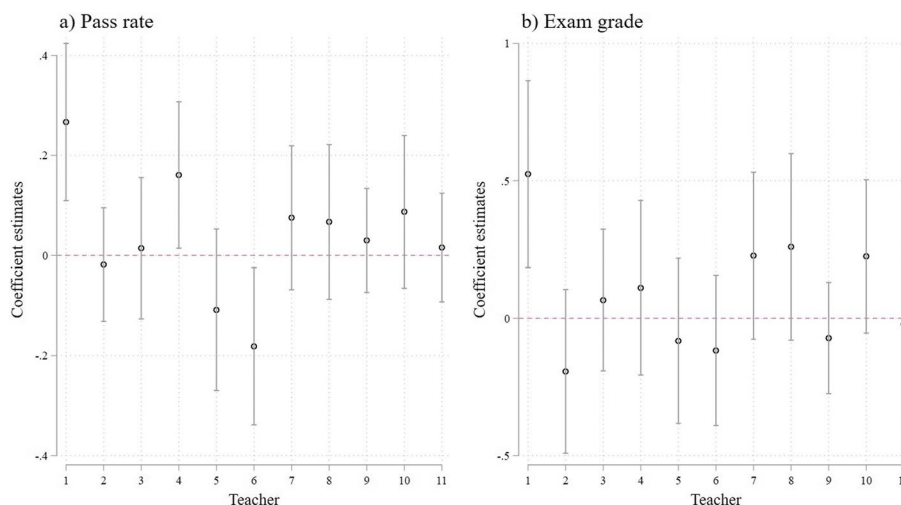

Fig. 2 Estimates of teacher specific treatment effects. Results from estimation of Eq. (2). Bars display 90% confidence intervals based on heteroscedasticity robust standard errors

Heterogeneity across teachers

To investigate the first research question, we present the estimates of the interaction terms of the model in Eq. (2) and their associated 90% heteroscedasticity robust confidence intervals visually in Fig. 2. The figure indicates that there is substantial variation in the treatment effect between teachers with the treatment effects varying from − 0.19 SDs to 0.52 SDs (exam grade) and − 18.2 to 26.7 percentage points (pass). When evaluating significance at a 10 percent level, two of the eleven teachers in our sample have positive treatment effects, one have negative treatment effects and the rest have insignificant treatment effects in the regressions with the students' pass rate as the outcome. For the exam grades, only Teacher 1 had a significant and positive treatment effect, while the treatment effect for all other teachers was too imprecisely measured for it to be statistically distinguished from zero.

To further explore the variations across teachers and hereby answer the second research question, we calculate the teacher effects separately by each treatment group based on the approach of mean correcting suggested by McCaffrey et al., (2012).

These mean corrected teacher effects are displayed in Fig. 3, where the teachers are sorted according to their effectiveness rank by treatment status. Several interesting insights arise from this figure. Perhaps the most striking one is that we observe some notable switches across treatment status, when looking at the ranking of teachers. There are two notable examples for the pass rate. First, observe that for Teacher 1 the change is from the position of being the relatively poorest teacher in the control group to the relatively best one in the treatment group. Second, for Teacher 6 the opposite is observed, as this teacher moves from being the second best teacher in the traditional classroom to being the relatively worst in the flipped classroom.

The pattern of rank reversal is only evident for some teachers, as Teacher 7, 9 and 11 are consistently at the middle of the teacher effectiveness distribution. When we look at the graphs with exam grade as the outcome, we again observe changes in the relative
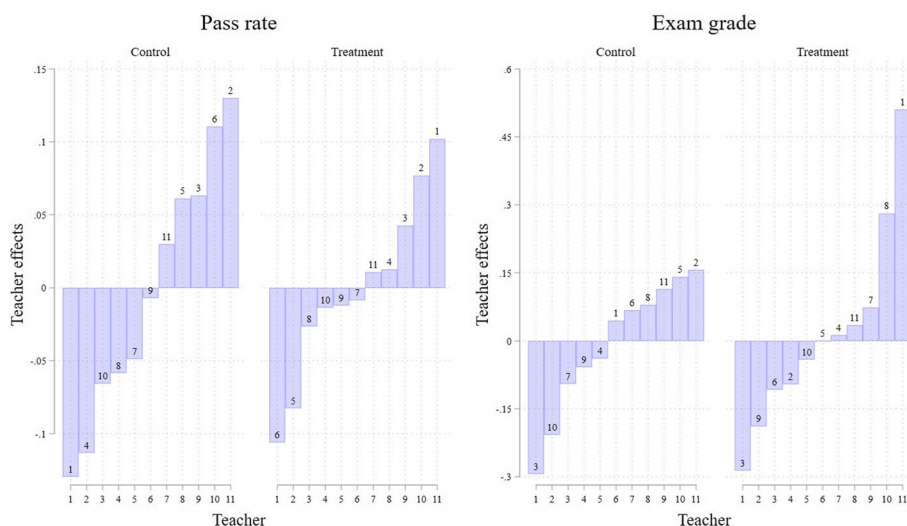

Fig. 3 Ranks of within-treatment teacher effects by control and treatment group. Based on method described in (McCaffrey et al. 2012). Includes baseline controls

teacher ranks, although none of the switches are as extreme as when we consider the pass rate outcomes. For example, Fig. 3 shows that Teacher 2, who is ranked as the best teacher in the control setting, is part of the low- to middle-ranked teachers in the flipped classroom setting. Moreover, the plot shows that while Teacher 1 by far has the highest teacher effectiveness in flipped classroom, he ranks in the middle of the distribution of teachers' effects on students' average exam grades in the control setting.

Given that class attendance is voluntary, one might wonder if the reason why we observe these switches in relative teacher ranks is due to selective tutorial class attendance among students: If students' attendance on average differs between flipped and traditional classrooms, this could explain the differences in teacher effectiveness across the two formats. Recall that the intervention was designed such that the time slots of the classes were flipped halfway through the semester. Therefore, we are not too concerned that any potential patterns in selective attendance is due to teachers leveraging their experiences with teaching the first class—whether it be the traditional or flipped classroom—to deliver a higher quality of teaching in the second class.

Since attendance is a post-treatment variable it would be a 'bad' control if included in the regression models. Instead, to get some descriptive insights on attendance, Fig. 4 shows average tutorial class attendance by teacher for all students in the full analytical sample (top panel) and for the subset of students who participated in at least one third of all tutorial classes (bottom panel). Class attendance for a given student is calculated as the share of tutorial classes in which this student showed up. We look at both of these
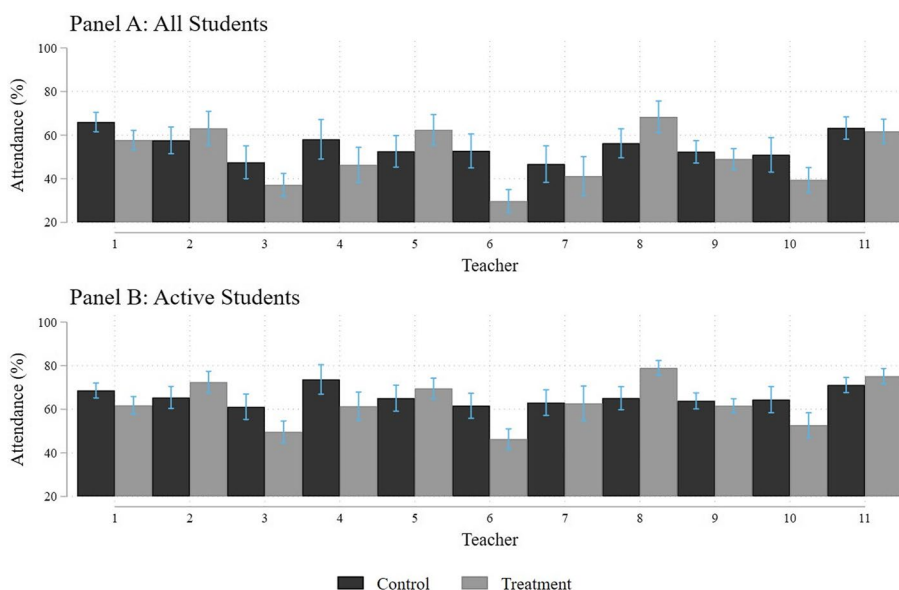

Fig. 4 Tutorial class attendance. Bars indicate 90% confidence intervals based on heteroscedasticity robust standard errors. Display raw differences without inclusion of controls. In Panel A the test of equality of means across treatment status for each teacher suggests a significant difference in the raw means for five teachers at a 10 percent significance level: Teacher 1 ($p = 0.034$), Teacher 3 ($p = 0.068$), Teacher 6 ($p = 0.000$), Teacher 8 ($p = 0.047$), and Teacher 10 ($p = 0.056$). In Panel B, the raw differences in attendance are significant for the same teachers as in Panel A and additionally also for Teacher 4: Teacher 1 ($p = 0.038$), Teacher 3 ($p = 0.018$), Teacher 4 ($p = 0.036$), Teacher 6 ($p = 0.002$), Teacher 8 ($p = 0.001$), and Teacher 10 ($p = 0.024$)

averages, because we want to see if students who never show up drive the overall mean attendance or if it is a general pattern for all students taught by the same teacher.

Figure 4 indicates that, on average, there is a higher attendance among the untreated students in traditional classrooms for both student populations. This tendency is particularly pronounced for some teachers, namely Teacher 3, 6, and 10. However, whereas Teacher 6 is one of the prominent examples of rank reversals, Teachers 3 and 10 do not exhibit the same pattern. Moreover, Teacher 1, who changes rank from bottom to top between the two pedagogical formats when considering the pass outcome, only has a small difference in attendance between the two different formats. The perhaps most important takeaway from Fig. 4 is that selective tutorial class attendance does not appear to be a main factor driving the observed teacher rank changes.

Overall, even though we only find few significant estimates of the interactions between teachers and treatment status, the rank analysis in this section does indicate that there might still be important teacher heterogeneity present. More specifically, the notable rank changes in Fig. 4 suggest that there is great variability in teachers' ability to reap the benefits of each of the traditional and flipped classroom format.

Though the effect of teachers is the most widely investigated classroom-level variable affecting student outcomes, the effect of peers has become another classroom factor receiving considerable attention from educational scholars (Sacerdote, 2011). Because both peers and teachers are defined at the classroom-level, one might worry that our results related to teachers are in fact driven by differences in the composition of a student's tutorial class peers. To examine whether this might be the case, we investigated the effect of including a measure of the average ability level of a student's tutorial class peers—the leave-self-out mean of high school GPA—in the estimations of the average treatment effect and in the computations of the relative teacher ranks. The results of these exercises are displayed in Appendix Table 6 and Appendix Fig. 6 and suggest no changes compared to the results without controls for peer ability levels.

Discussion and implications

The study's results suggest modest heterogeneities in the average effect of flipped classroom across teachers, but that the effectiveness ranks of teachers vary notably across the two different teaching formats. Importantly, these findings do not appear to be driven by peer composition. The number of teacher rank changes is noteworthy, as the estimates are obtained from a very controlled setting where the teachers had explicit instruction on how to teach the flipped classroom condition. In addition, all teachers are similar on basic observable characteristics; all except one are males, most have extensive experience, they are all part time teachers and are roughly around the same age. This corresponds to the insights arising from the literature on total teacher effectiveness (e.g. Kane & Staiger, 2008; Rivkin et al., 2005; Rockoff, 2004) which suggests that unobservable teacher characteristics are more important than observable ones. This could suggest that the changes in teacher ranks are more likely to stem from unobservable characteristics such as personality, teaching style or attitudes towards new teaching formats. The results are limited by the fact that the study only includes eleven teachers, which means that

going one step further and correlating the teacher effects with observed characteristics or attempting to estimate total teacher effectiveness in each format is out of the scope for this paper. Instead, we suggest this as a potential subject for future research.

A potential concern related to the notion of differences in teacher abilities across the treatment and control setting is whether the observed rank reversals are simply due to unobserved idiosyncratic variance. Although we cannot rule out the presence of any important unobservables, we do consider it unlikely that systematic variation in student characteristics should explain the differences in teacher ranks given the randomization and balance of observables in the two pedagogical formats cf. Table 2.

The findings of this study have important implications for practice. The increasing use of technology-supported teaching and learning formats places responsibility for managing the educational change process on teachers and institutions as mentioned by Bruggeman et al. (2021). Teachers are central to this process and as our findings show, their ability to transfer their teaching competencies between traditional classroom teaching and flipped classroom (and vice versa) varies substantially across teachers. In the present study, the preparation of the intervention involved teachers participating in a couple of pedagogical workshops, which, given the findings, might not have been enough to facilitate successful implementation of the flipped classroom format. This highlights that teachers' attributes and skills are critical and should be identified and developed, if educational policy makers are to reap the potential benefits that flipped classroom has to offer student learning (see for example Strelan et al., 2020). The expert interviews by Bruggeman et al. (2021) provide relevant knowledge on attributes for (mal)adaptation of blended learning more broadly and future studies may build on this to systematically investigate and test different teacher attributes to generate knowledge about faculty development activities that can facilitate the changes to flipped classroom. This, in turn, could support teachers as well as institutions in the ongoing organizational change process to implement flipped classroom in higher education.

Limitations

While our study indicates the importance of teachers for success of flipped classroom inspired teaching, some limitations need to be addressed. First, our study is based on data from a single university. While there are many similarities between higher education institutions they may also differ substantially according to regional/national rules and regulations, and we hope to have inspired others to explore the role of teachers in other contexts. Moreover, the study contains a small number of teachers which also limit its generalizability.

Conclusion

This study complements recent literature on the effects of flipped classroom by investigating heterogeneous treatment effects across teachers. We utilize two iterations of a randomized flipped classroom intervention and find a positive yet statistically insignificant effect of flipped classroom on both pass rate and final exam grades. Focusing on the effect of different teachers, we see few cases of significant teacher-treatment

heterogeneity. However, we do find substantial shifts in the ranks of teacher effectiveness between the traditional and flipped classroom classes, suggesting that the best teacher in a traditional teaching environment is not necessarily the best teacher in a flipped classroom environment. These results show that even in a highly controlled environment (such as a field experiment) teachers play a role for the effectiveness of flipped classroom. Accordingly, more research is needed on what constitutes a good teacher in a flipped classroom environment, as this appears to differ from a traditional setting.

This leads to some final considerations for future developments for research. While research on flipped classroom is rapidly increasing, the literature is characterized by few studies that focus on the effects of teachers. In this article, we report from a systematic study of teacher effects in flipped classroom, however a number of questions are still unanswered and need to be addressed. First, due to the small number of teachers in our sample, the findings can only be indicative, and we encourage others to replicate it with larger numbers of teachers but also in other contexts as regional regulatory and cultural differences may have significant impact on teachers' implementation of the format. Moreover, questions that focus on a particular issue in relation to teachers' use of the format would be relevant to answer, including if certain active learning techniques are easier to use for some teachers than others or what support and training teachers benefit from when engaging in this teaching format.

Also, in our setting, the RCT was initiated by the education institution and not by the course responsible or the teachers themselves. Our findings therefore report on effects by teachers who have not chosen to engage in this form of teaching. Future studies could focus on teachers' motivation and other personal characteristics including personality characteristics and teaching preferences to better understand how to prepare teachers for flipped classroom. This may include hypothesis testing of, for example, if teachers who have a 'learner-focus' rather than a 'subject-focus' (Kolb & Kolb, 2014) are more likely to succeed in increasing student learning outcome or if teachers scoring high on extroversion (McCrae & John, 1992) feel more comfortable teaching in a flipped classroom format. This would provide relevant information for faculty development and how to best support teachers when teaching in a flipped classroom format.

Finally, the relationship between teachers and students in flipped classroom is a relevant topic to be researched in more depth. For example, it would be useful to know more about how teachers can scaffold the learning process in flipped classroom to help students better engage in and gain the benefits of the method. This could help to solve issues of student reluctance to participate. Regarding another often mentioned issue—students lack of preparation for flipped classroom in-class sessions, further research on how teachers address this challenge successfully could be useful for increasing the impact of the format.

Appendix
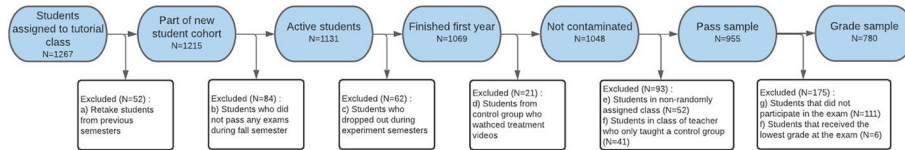
See Figures 5, 6 and Tables 5, 6.
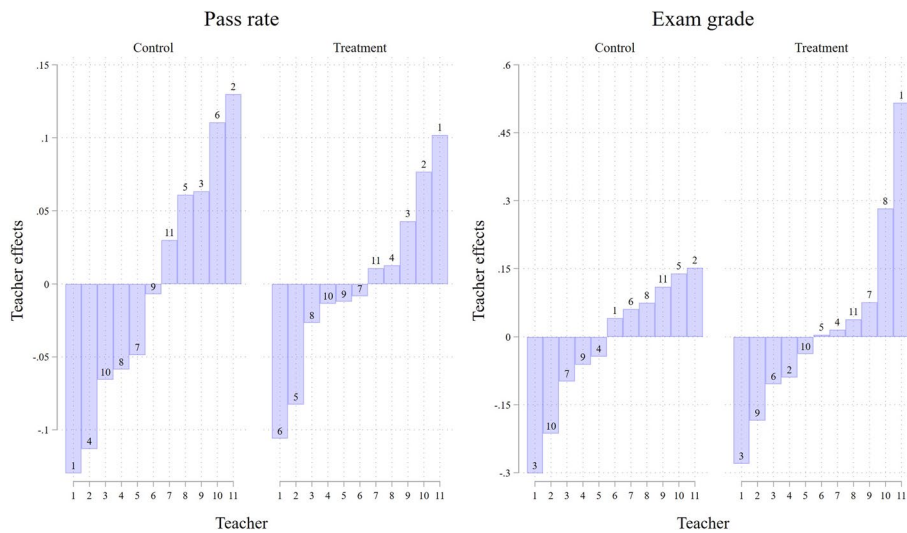

Fig. 5 Data cleaning

Fig. 6 Ranks of within-treatment teacher effects by control and treatment group with peer control. Based on method described in McCaffrey et al. (2012). Includes baseline controls and leave-self-out mean of high school GPA computed at the tutorial class level

Table 5 Summary of previous literature mentioning teacher effects in flipped classroom in higher education

Authors	Scope of study	Scope with respect to impact of teachers	Methodology	Type of analysis
Kim et al. (2014)	Design principles for flipped classroom derived from three case studies	The role of teachers is changed in FC to become that of initiator, facilitator	Mixed methods: Surveys, interviews with students, instructor reflections, document analysis. Descriptive statistics, t-tests, correlation coefficient analysis	Three classes (engineering, sociology, humanities). 115 students. 3 instructors each with a unique implementation of flipped classroom
Karabulut-Ilgu et al. (2018)	Systematic review of flipped learning in engineering education	Teachers' role is discussed. Biggest challenge: heavy workload both before and during class. Student resistance	Descriptive statistics on number of articles by year, number of articles by journal, theoretical frameworks applied, effect evaluation results, benefits and challenges	Systematic review. 62 articles from between 2000 and 2015. Details about articles selection provided. Summaries to extract knowledge
Akçayır and Akçayır (2018)	Systematic review of flipped classroom. Focus on advantages, challenges and chosen activities	Teachers' new role. Challenges for teachers: time consuming, difficult to manage and difficult to plan activities	Descriptive statistics on number of articles by year, learner types (Higher education 80%). Categories for advantages, challenges, and activities	Systematic review. 71 articles from between 1980–2016. Summaries and categorizations to extract knowledge
DeLozier and Rhodes (2017)	Review article. Ideas and recommendations for practice. Focus on various elements of flipped classroom out-of and in-class	Teacher' role in combination with different activities. Mentions differences in instructor implementation	Discussions on out-of-class lectures, in-class activities, pair-and-share activities, student presentations, Active learning with the flipped classroom	Review article. References from 1976 to 2015. In total 90 references
Hwang et al. (2015)	A theoretical study on seamless flipped learning (SFL)	Elaboration on the new role of teachers as facilitators. Introduces five success criteria for teachers' implementation of SFL	To derive recommendations based on theoretical considerations	Introduction of SFL model. Not exclusively for higher education
Ghadiri (2014)	Case study reporting from developing and implementing blended learning on one course within STEM	Approach very demanding for faculty. New role. Discusses what teachers can learn from various elements of blended learning and flipped classroom	Descriptive comparisons of grade and pass rate outcomes from own implementation of BL. Histograms. ANOVA. Linear multiple regression	Cohort from 2012 to 2014. An intro course in 'Circuit Analysis'. Three classes of 50, 78, 75 students respectively
Zou et al. (2020)	Systematic review of flipped language classroom	Teachers' perceptions are mentioned. A section on implications for teachers	Classifications. Focus on course elements, learning tools (pre-class, in-class and after-class). Evaluation methods and effects of flipped classroom	Systematic review. Based on 34 articles from 2015 to 2019 in 14 different journals. Mostly higher education

Some articles use the term 'teachers' other 'instructors' others again 'educators'. We apply the terminology of each paper in this table

Table 6 Average treatment effects with peer effect controls

	Pass rate		Exam grade	
	(1)	(2)	(3)	(4)
Treatment	0.037	0.037	0.059	0.059
	[0.306]	[0.332]	[0.330]	[0.342]
Experiment year	0.023	0.026	− 0.173**	− 0.169*
	[0.596]	[0.622]	[0.026]	[0.058]
GPA Fall	0.054***	0.054***	0.252***	0.252***
	[0.000]	[0.000]	[0.000]	[0.000]
High School GPA	0.001	0.002	0.072***	0.073***
	[1.000]	[0.926]	[0.002]	[0.002]
Female	0.001	0.000	− 0.152**	− 0.152**
	[0.958]	[0.990]	[0.014]	[0.012]
Age	− 0.009	− 0.009	0.015	0.015
	[0.296]	[0.280]	[0.292]	[0.302]
Retake microeconomics exam	− 0.238***	− 0.238***	− 0.045	− 0.045
	[0.000]	[0.000]	[0.486]	[0.492]
Mean of peer high school GPA		0.098		0.072
		[0.368]		[0.570]
Observations	933	933	763	763
R-squared	0.282	0.283	0.516	0.516
Teacher fixed effects	Yes	Yes	Yes	Yes

p-values in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1. Uses wild cluster bootstrap standard errors. Clustered at the class level with B = 1000

Declarations

Ethics approval and consent to participate
Ethical approval was received from the institutional ethics review committee.

Competing interests
The authors declare that they have no competing interests.

References

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

Bergmann, J., & Sams, A. (2012). Flip Your Classroom: Reach Every Student in Every Class Every Day. International Society for Technology in Education.

Brewer, R., & Movahedazarhouligh, S. (2018). Successful stories and conflicts: A literature review on the effectiveness of flipped learning in higher education. *Journal of Computer Assisted Learning, 34*(4), 409–416. https://doi.org/10.1111/jcal.12250

Bruggeman, B., Tondeur, J., Struyven, K., Pynoo, B., Garone, A., & Vanslambrouck, S. (2021). Experts speaking: Crucial teacher attributes for implementing blended learning in higher education. *The Internet and Higher Education, 48*, 100772. https://doi.org/10.1016/j.iheduc.2020.100772

Calimeris, L., & Sauer, K. M. (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

Carrell, S. E., & West, J. E. (2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy, 118*(3), 409–432. https://doi.org/10.1086/653808

Chen, F., Lui, A. M., & Martinelli, S. M. (2017). A systematic review of the effectiveness of flipped classrooms in medical education. *Medical Education, 51*(6), 585–597. https://doi.org/10.1111/medu.13272

Chen Hsieh, J. S., Huang, Y.-M., & Wu, W.-C.V. (2017). Technological acceptance of LINE in flipped EFL oral training. *Computers in Human Behavior, 70*, 178–190. https://doi.org/10.1016/j.chb.2016.12.066

Colin Cameron, A., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources, 50*(2), 317–372. https://doi.org/10.3368/jhr.50.2.317

DeLozier, S. J., & Rhodes, M. G. (2017). Flipped classrooms: A review of key ideas and recommendations for practice. *Educational Psychology Review, 29*(1), 141–151. https://doi.org/10.1007/s10648-015-9356-9

Ficano, C. K. C. (2019). Identifying differential benefits from a flipped-group pedagogy in introductory microeconomics. *International Review of Economics Education, 30*, 100143. https://doi.org/10.1016/j.iree.2018.07.002

Ghadiri, K. (2014). Developing and implementing effective instructional strategems in STEM. *ASEE Annual Conference & Exposition, 2014*, 24–391.

Hanushek, E. A., & Rivkin, S. G. (2006). Chapter 18 Teacher Quality. I *Handbook of the Economics of Education* (Bd. 2, s. 1051–1078). Elsevier. https://doi.org/10.1016/S1574-0692(06)02018-6

Hwang, G.-J., Lai, C.-L., & Wang, S.-Y. (2015). Seamless flipped learning: A mobile technology-enhanced flipped classroom with effective learning strategies. *Journal of Computers in Education, 2*(4), 449–473. https://doi.org/10.1007/s40692-015-0043-0

Kane, T., & Staiger, D. (2008). Estimating teacher impacts on student achievement: An experimental evaluation (Nr. w14607; s. w14607). *National Bureau of Economic Research*. https://doi.org/10.3386/w14607

Karabulut-Ilgu, A., Jaramillo Cherrez, N., & Jahren, C. T. (2018). A systematic review of research on the flipped learning method in engineering education. *British Journal of Educational Technology, 49*(3), 398–411. https://doi.org/10.1111/bjet.12548

Kim, M. K., Kim, S. M., Khera, O., & Getman, J. (2014). The experience of three flipped classrooms in an urban university: An exploration of design principles. *The Internet and Higher Education, 22*, 37–50. https://doi.org/10.1016/j.iheduc.2014.04.003

Kolb, A. Y., Kolb, D. A., Passarelli, A., & Sharma, G. (2014). On becoming an experiential educator: The educator role profile. *Simulation & Gaming, 45*(2), 204–234. https://doi.org/10.1177/1046878114534383

Lai, C.-L., & Hwang, G.-J. (2016). A self-regulated flipped classroom approach to improving students' learning performance in a mathematics course. *Computers & Education, 100*, 126–140. https://doi.org/10.1016/j.compedu.2016.05.006

Lai, H.-M. (2021). Understanding what determines university students' behavioral engagement in a group-based flipped learning context. *Computers & Education, 173*, 104290. https://doi.org/10.1016/j.compedu.2021.104290

Lai, H.-M., Hsieh, P.-J., Uden, L., & Yang, C.-H. (2021). A multilevel investigation of factors influencing university students' behavioral engagement in flipped classrooms. *Computers & Education, 175*, 104318. https://doi.org/10.1016/j.compedu.2021.104318

Lo, C. K., & Hew, K. F. (2019). The impact of flipped classrooms on student achievement in engineering education: A meta-analysis of 10 years of research. *Journal of Engineering Education, 108*(4), 523–546. https://doi.org/10.1002/jee.20293

Love, B., Hodge, A., Grandgenett, N., & Swift, A. W. (2014). Student learning and perceptions in a flipped linear algebra course. *International Journal of Mathematical Education in Science and Technology, 45*(3), 317–324. https://doi.org/10.1080/0020739X.2013.822582

McCaffrey, D. F., Lockwood, J. R., Mihaly, K., & Sass, T. R. (2012). A review of stata commands for fixed-effects estimation in normal linear models. *The Stata Journal: Promoting Communications on Statistics and Stata, 12*(3), 406–432. https://doi.org/10.1177/1536867X1201200305

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*(2), 175–215.

McLaughlin, J. E., Roth, M. T., Glatt, D. M., Gharkholonarehe, N., Davidson, C. A., Griffin, L. M., Esserman, D. A., & Mumper, R. J. (2014). The flipped classroom: A course redesign to foster learning and engagement in a health professions school. *Academic Medicine, 89*(2), 236–243. https://doi.org/10.1097/ACM.0000000000000086

Mingorance, E., Granda, V., Rojas, R., & Alemany, A. (2019). Flipped classroom to improve university student centered learning and academic performance. *Social Sciences, 8*(11), 315. https://doi.org/10.3390/socsci8110315

Nielsen, P. L., Bean, N. W., & Larsen, R. A. A. (2018). The impact of a flipped classroom model of learning on a large undergraduate statistics class. *Statistics Education Research Journal, 17*(1), 121–140. https://doi.org/10.52041/serj.v17i1.179

Nouri, J. (2016). The flipped classroom: For active, effective and increased learning—especially for low achievers. *International Journal of Educational Technology in Higher Education, 13*(1), 33. https://doi.org/10.1186/s41239-016-0032-z

O'Flaherty, J., & Phillips, C. (2015). The use of flipped classrooms in higher education: A scoping review. *The Internet and Higher Education, 25*, 85–95. https://doi.org/10.1016/j.iheduc.2015.02.002

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–458. https://doi.org/10.1111/j.1468-0262.2005.00584.x

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247–252. https://doi.org/10.1257/0002828041302244

Ryan, M. D., & Reid, S. A. (2016). Impact of the flipped classroom on student performance and retention: A parallel controlled study in general chemistry. *Journal of Chemical Education, 93*(1), 13–23. https://doi.org/10.1021/acs.jchemed.5b00717

Sacerdote, B. (2011). Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? I *Handbook of the Economics of Education* (Bd. 3, s. 249–277). Elsevier. https://doi.org/10.1016/B978-0-444-53429-3.00004-1

Setren, E., Greenberg, K., Moore, O., & Yankovich, M. (2021). Effects of flipped classroom instruction: Evidence from a randomized trial. *Education Finance and Policy, 16*(3), 363. https://doi.org/10.1162/edfp_a_00314

Strelan, P., Osborn, A., & Palmer, E. (2020). The flipped classroom: A meta-analysis of effects on student performance across disciplines and education levels. *Educational Research Review, 30*, 100314. https://doi.org/10.1016/j.edurev.2020.100314

Sun, Z., Xie, K., & Anderman, L. H. (2018). The role of self-regulated learning in students' success in flipped undergraduate math courses. *The Internet and Higher Education, 36*, 41–53. https://doi.org/10.1016/j.iheduc.2017.09.003

van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis. *Educational Research Review, 28*, 100281. https://doi.org/10.1016/j.edurev.2019.05.003

Wozny, N., Balser, C., & Ives, D. (2018). Evaluating the flipped classroom: A randomized controlled trial. *The Journal of Economic Education, 49*(2), 115–129. https://doi.org/10.1080/00220485.2018.1438860

Zou, D., Luo, S., Xie, H., & Hwang, G.-J. (2020). A systematic review of research on flipped language classrooms: Theoretical foundations, learning activities, tools, research topics and findings. *Computer Assisted Language Learning, 35*(8), 1811–1837. https://doi.org/10.1080/09588221.2020.1839502

Publisher's Note