OSIPP Discussion Paper : DP-2025-E-001

The Effects of Flipped Classroom in Higher Education on Learning Outcomes and the Heterogeneity by Group Member Characteristics

May 12, 2025

Yasukazu Ichino*
Professor, Economics Department, Ritsumeikan University

Toru Kawai
Associate Professor, Sport and Health Science Department, Ritsumeikan University

Mifuyu Kira
PhD.student, Osaka School of International Public Policy, The University of Osaka

Mai Seki
Professor, Faculty of International Social Sciences, Gakushuin University

【JEL Classification】 A22, I23, C93
【Keywords】 flipped classroom; randomized experiment; group effects
【Abstract】While peer effects in higher education, particularly through random roommate assignments, have been extensively studied, research on in-class peer effects, especially in group work settings, remains limited. Simultaneously, flipped classrooms are gaining increasing attention in higher education as a popular form of active learning, where students watch lecture videos before class and engage in group discussions or assignments during class. This study aims to estimate the heterogeneous impact of flipped classrooms on college students' learning outcomes by examining how these effects vary according to the characteristics of group members (i.e., math and English placement test scores, learning motivation, and attitude toward group learning). This study uses random assignment of first-year college students to sections of an introductory economics course and random assignment of group members within the flipped classroom (totaling n=3749), to examine the impact of the flipped classroom teaching method. The study finds that the presence of highly motivated group members significantly improves students' learning outcomes. The results provide suggestive evidence that peer characteristics could positively influence learning outcomes in in-class activities within flipped classrooms.

*Corresponding Author. (yir10042@fc.ritsumei.ac.jp)

The Effects of Flipped Classroom in Higher Education on Learning Outcomes and the Heterogeneity by Group Member Characteristics

Yasukazu Ichino, Toru Kawai, Mifuyu Kira, and Mai Seki

May 12, 2025

While peer effects in higher education, particularly through random roommate assignments, have been extensively studied, research on in-class peer effects, especially in group work settings, remains limited. Simultaneously, flipped classrooms are gaining increasing attention in higher education as a popular form of active learning, where students watch lecture videos before class and engage in group discussions or assignments during class. This study aims to estimate the heterogeneous impact of flipped classrooms on college students' learning outcomes by examining how these effects vary according to the characteristics of group members (i.e., math and English placement test scores, learning motivation, and attitude toward group learning). This study uses random assignment of first-year college students to sections of an introductory economics course and random assignment of group members within the flipped classroom (totaling n=3749), to examine the impact of the flipped classroom teaching method. The study finds that the presence of highly motivated group members significantly improves students' learning outcomes. The results provide suggestive evidence that peer characteristics could positively influence learning outcomes in in-class activities within flipped classrooms.

JEL code: A22, I23, C93
Keywords: flipped classroom; randomized experiment; group effects

Author Affiliations and Acknowledgments

Yasukazu Ichino (corresponding author: yir10042@fc.ritsumei.ac.jp), Economics Department, Ritsumeikan University; Toru Kawai, Sport and Health Science Department, Ritsumeikan University; Mifuyu Kira Osaka School of International Public Policy, The University of Osaka; Mai Seki, Faculty of International Social Sciences, Gakushuin University. We thank the participants of Tokyo Labor Economics Workshop at the University of Tokyo, Tokyo Keizai University's Applied Microeconomics Seminar, Tohoku University Modern Economics Seminar, and the Kyoto Applied Microeconomics Seminar at Doshisha University for their helpful comments on earlier manuscripts. We have extensively revised the draft to reflect the constructive comments and discussions from AASLE 2024 conference. The opinions expressed in this paper are those of the authors and do not reflect the views of affiliated organizations. We would like to thank Editage (www.editage.jp) for English language editing. This study was supported by a Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (C) No. 22K01527. IRB is approved by the Ritsumeikan University IRB committee No. Kinugasa-Jin-2021-112.

Declaration of competing interests and Data availability statement

The authors declare no relevant financial or non-financial competing interests. Owing to the nature of this research, the participating universities' permission to share the data is needed upon each request by a researcher.

1 Introduction

Peer effects are a significant area of interest among education stakeholders, including university faculties. Studies on peer effects in higher education that use random roommate assignments are well-established and have shown that college roommates' behaviors significantly influence learning outcomes (Sacerdote, 2001; Zimmerman, 2003; Kremer and Levy, 2008; Carrell et al., 2009). However, few studies have examined in-class peer effects in the context of higher education (Hong and Lee, 2017). Moreover, Kimbrough et al. (2022) is the only study to explore peer effects in group work settings in higher education, while more studies have addressed this topic at the K-12 level (Li et al., 2014; Dong et al., 2023; Wang et al., 2024). This study aims to explore how group member characteristics in in-class discussion groups influence students' learning outcomes in higher education. This area has received limited attention and requires further investigation to enhance college students' learning experiences. Specifically, we focus on the role of non-cognitive characteristics, as they are known to influence cognitive development and long-term outcomes(Heckman and Rubinstein, 2001; Heckman et al., 2006; Cunha et al., 2006; Cunha and Heckman, 2008; Heckman and Kautz, 2012).

We examine the effects of in-class group discussions within a flipped classroom setting. Flipped classrooms, a form of active learning, have become increasingly popular as a means of incorporating lecture videos and in-class activities. In a typical flipped classroom, students watch lecture videos before class and engage in assignments or group discussions during class. Unlike traditional in-person lectures, students can watch videos at their own pace, removing the need to remain seated and focused for a fixed period. During in-class discussions, students receive real-time feedback from peers, teaching assistants, and instructors as they work on assignments and engage in conversations. While most studies report the positive effects of flipped classrooms on student learning outcomes, such as im-

proved exam scores, some have found insignificant effects (Caviglia-Harris, 2016; Balaban et al., 2016; Olitsky and Cosgrove, 2016; Swoboda and Feiler, 2016; Calimeris and Sauer, 2015; Wozny et al., 2018; Craft and Linask, 2020). In the replication of our previous study using the latest data, we have found that approximately 70 percent of the positive effects of flipped classrooms on students' learning stem from viewing lecture videos, with the remaining effects attributed to others, including in-class activities.[1] This raises the question of whether group discussions and peer learning play any role in in-class activities and if peer characteristics contribute to enhancing the effects of flipped classrooms.

In this study, we investigate the mechanism behind group-learning effects by examining the heterogeneity of impacts based on group member characteristics (i.e., attitude toward group learning, learning motivation, placement test scores) in flipped classrooms. Our results indicate that the presence of highly motivated group members significantly improves a student's learning outcomes in flipped classrooms, while group members' placement test scores (i.e., initial academic abilities) do not have a significant effect. The identification of the flipped classroom effect is based on the random assignment of college students to classes by department administrators, and the identification of the heterogeneous effect of group member characteristics relies on the random assignment of groups by the instructor.

The remainder of this paper is organized as follows. Section 2 explains the research design, and Section 3 describes the data. Section 4 introduces the empirical specifications. Section 5 presents the results, including a mediation analysis and discussion. Section 6 concludes.

[1] Appendix A presents the details of the mediation analysis results using 2022 and 2023 data to replicate Ichino et al. (2025). The mediation effect through watching lecture videos is 72 percent (=0.174/0.243) while the direct effect is 28 percent.

2 Research Design

2.1 Experimental design

We conducted our experiment in the Spring semester of 2022 and 2023 in "Introduction to Economics," a required course for first-year students in the Department of Economics of a private university in Japan. Introduction to Economics consists of four sections. The administrative office of the economics department assigned students to these four sections, A, B, C, and D, based on the information from their college entrance examsThis assignment aims to ensure that students' initial abilities are roughly equal across all sections. Students assigned to Sections A and B did not participate in the experiment; those assigned to Sections C and D did.

In each section, the class met every Monday from 9:00 am to 10:30 am for a total of 14 lessons. Each section was divided into two components, an introduction to micro and macroeconomics and an introduction to heterodox economics, with seven lessons each. The experiment was conducted in the micro and macroeconomics component taught by the same instructor in the first seven lessons in Section C and the second seven lessons in Section D (see Figure 1).

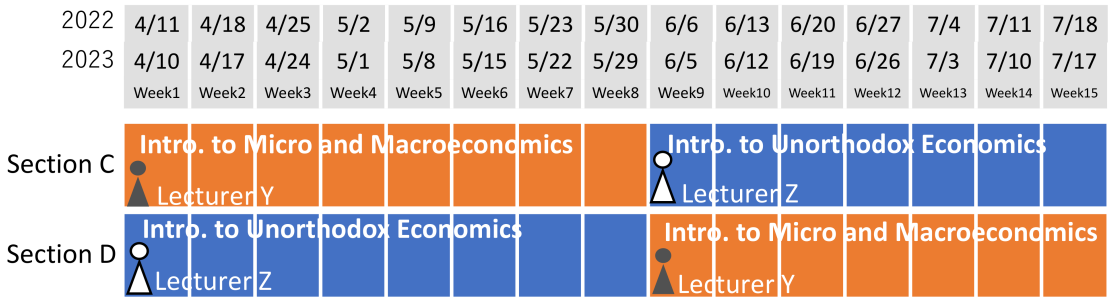

Figure 1: Introduction to Economics in 2022 and 2023

To ensure equal learning opportunities, we had to design the experiment so that students would experience both the flipped classroom and traditional lectures, with only timing

differences. Specifically, among the seven lessons, lessons 1–4 were taught in the traditional lecture format for Section C and in the flipped classroom format for Section D, while lessons 5–7 were taught in the flipped classroom format for Section C and in the traditional lecture format for Section D. All lessons were taught in in-person classes. The quiz questions for the same lesson were different in Sections C and D, although the level of difficulty was identical. This was to avoid cheating across sections by utilizing the timing difference of the lessons taught. The differences in class structure between the traditional lecture and the flipped classroom in our experiment are summarized in Figure 2.

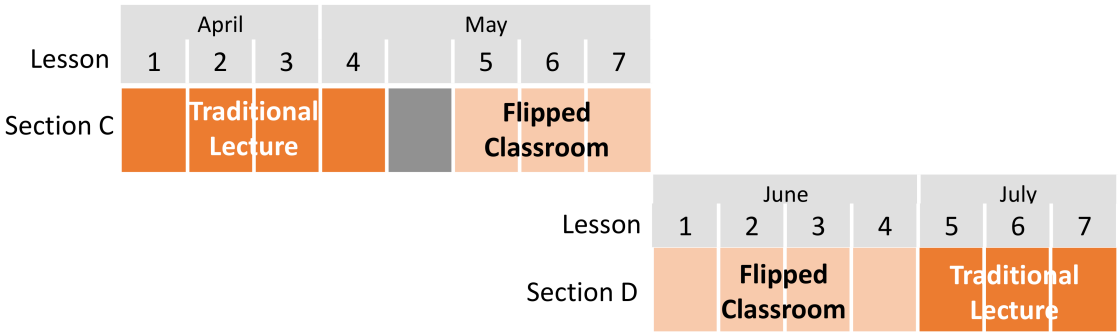

Figure 2: Research Design in 2022 and 2023

2.2 Traditional lecture vs. flipped classroom

In the traditional lecture format, students were asked to read the assigned pages of the textbook and supplementary readings before class. During class, the instructor gave lectures with lecture content displayed on slides. The instructor sometimes provided additional notes and comments using blackboards or by making handwritten annotations on lecture slides. After class, the students were required to submit a question about the lecture to the course website. Additionally, students were required to work on practice problems on their own after class and submit their answers to the course website. The students were notified that problems similar to these practice problems would be asked in the in-class

quiz, administered during the first 10 minutes of the class time following the lecture (one week later). The quiz comprised five multiple-choice questions.

In the flipped classroom format, students were asked to watch video lectures and read assigned pages of the textbook and supplementary readings before class. The video lectures comprised the same slides used in the traditional lectures. In the video lectures, the instructor annotated the slides using a pen tool, as in traditional lectures. In the bottom-right corner of each slide, a bust-shot video of the instructor speaking, sometimes with gestures, was shown. Students were required to submit questions about the video lectures as homework assignments. During class, students worked on practice problems in groups of six students, with the members of the groups randomly assigned by the instructor. The assigned groups were the same for all flipped classes. The practice problems were identical to those given in the traditional lecture format. During class, the instructor and four to five teaching assistants (TAs) walked around the classroom to help the groups answer their questions. If assistance was requested by the students, the TAs and instructor would provide the answers to the practice problems and explain how they were obtained. In the flipped classrooms, as in the traditional lectures, the students were required to submit their answers to the practice problems to the course website. In the flipped classroom format, the quizzes were given in the last 10 minutes of the class.[2]

In summary, in this study, the treatment effects of the flipped classroom , in contrast to the traditional classroom, arose from stronger incentives to watch on-demand video lectures (e.g., students rarely watch the video in advance in traditional classrooms but do watch the

[2]In the flipped classroom format, the quiz was given immediately after the content was covered, whereas in the traditional lecture, the quiz was given one week later. However, in the traditional lecture format, students could work on practice problems any time they preferred. They were even able to review the practice problems immediately before taking the quiz. In this sense, the time difference between working on practice problems and taking the quiz is almost the same for the flipped classroom and traditional lecture formats.

video in the flipped classroom) and in-class group discussions (e.g., peer effects and real-time support from TAs or instructors). All course resources prepared by the instructor, such as supplementary readings, slides, video lectures, and practice problems and their answer keys, were identical across each format. Additionally, because we allowed students to work on exercises in assigned groups within the flipped classrooms, we can examine the potential heterogeneous effects of flipped classrooms on student learning based on group member characteristics.

2.3 Measurements of learning outcomes

In this experiment, the students' learning outcomes were quizzes and final exam scores. Because the quizzes were given within a week of the lecture, the quiz scores were considered short-term learning outcomes. We obtained quiz scores for each student in each lesson. Since the final exam was taken at the end of the semester, exam scores were considered long-term learning outcomes, following the literature. However, as analyzed in Ichino et al. (2025), because flipped classrooms had little impact on final exam scores, we use only the quiz scores as a measurement of learning outcomes in this paper.

The grade for the course was assigned as follows: each component contributed 50% of the final grade. Of the 50% of the grades assigned to the introduction to micro and macroeconomics, 20% were for the quizzes, 10% for question submissions, 5% for answers to the practice problems, and 15% for the final exam. Given this grading rule, we believe students had a strong enough incentive to score well on quizzes.

3 Data

In 2022, there were 185 students in Section C and 179 students in Section D, with 180 and 158 students who consented to provide data for the study, respectively. In 2023, Section

C had 209 students, and Section D had 210 students, with 204 and 173 participating in the study, respectively. There are 5005 possible student-lesson observations for the quiz, with 715 students (338 and 377 from each year) and seven lessons. However, some students missed some quizzes, baseline placement tests, or surveys; therefore, the actual number of observations for quiz scores is 3749.

In addition to the quiz scores, we collected variables for the characteristics of the students such as gender, academic abilities, and noncognitive traits. Academic abilities are measured by scores on the math and English placement tests, completed at the beginning of the semester. Regarding noncognitive characteristics, we assessed learning motivation and attitude toward learning from survey questions. Learning motivation is based on a single survey question (i.e., How motivated are you to study at university?), and the outcome is measured on the 4-point Likert scale – (1) Motivated; (2) Slightly motivated; (3) Not very motivated; and (4) Not motivated. Attitude toward group learning is based on eight PISA 2015 questions (e.g., A team makes better decisions than one person; It's fun to work with classmates; Listen carefully to others). The outcome is measured on a 5-point Likert scale – (1) Strongly Disagree; (2) Disagree; (3) Neither Agree nor Disagree; (4) Agree; and (5) Strongly Agree.

Table 1 provides summary statistics on the basic student characteristics (i.e., gender, placement test scores) and student learning outcomes (i.e., quiz scores). In the regression analysis, the quiz scores for each lesson are standardized to a mean of zero and a standard deviation of one. In other words, the mean and standard deviation of quiz scores in each lesson are calculated over the two sections and applied to standardize the quiz scores of the students in each lesson.

To examine the possible heterogeneous effects of flipped classrooms on student learning by group member characteristics, we construct the group characteristic variables for each student from the following four student characteristic variables: math placement test

	Ave.	Std. dev.	Min	Max	Obs.
Student characteristics					
Gender (Woman = 1)	0.244	0.430	0	1	685
Math placement test score (Full score = 42)	23.725	6.627	2	40	685
English placement test score (Full score = 1000)	533.402	110.671	76	782	685
Learning outcome					
Quiz score (7 times)	3.150	1.321	0	5	3749
Assigned group characteristics					
Math Placement Test Score	1.071	0.631	-1.480	2.535	685
English Placement Test Score	0.962	0.518	-1.222	3.019	685
Attitude toward group learning	4.436	0.363	3.375	5.000	685
Learning motivation	3.987	0.114	3.000	4.000	685
Group Characteristics					
Math Placement Test Score	1.010	0.681	-1.926	2.535	3749
English Placement Test Score	0.928	0.556	-2.962	3.019	3749
Attitude toward group learning	4.393	0.387	2.625	5.000	3749
Learning motivation	3.971	0.168	3.000	4.000	3749

Table 1: Summary statistics

score, English placement test score, learning motivation, and attitude toward group learning. Specifically, the group characteristic of student i regarding, for example, the math test score, is defined as the highest math score among the members of the student i's group excluding student i. The summary statistics for such group characteristic variables are shown in the lines indicated by "Assigned group characteristics" in Table 1. Furthermore, another type of group characteristic variable is constructed; that is, the variable for attending group members, which is defined as the highest math score among the attending members of the student i's group excluding student i. The summary statistics for group characteristic variables for attending group members are shown in the lines indicated by "Group Characteristics" in Table 1. It should be noted that the assigned group characteristic variables

are constant over the lessons and thus treated as exogenous since it is determined before the experiment by the researcher; however, the group characteristic variables for attending members, vary over the lessons and should be treated as endogenous. In Table 1, math and English placement test scores are standardized.

	Section D Mean	Section C Mean	Difference	Observations
Math Placement Test Score	23.36	24.02	-0.660	685
	(6.72)	(6.54)	[0.194]	
English Placement Test Score	528.46	537.69	-9.23	685
	(111.37)	(110.04)	[0.277]	
Women	0.255	0.234	0.020	685
	(0.436)	(0.424)	[0.536]	

Standard deviations in parentheses for Treatment and Control Means in 2022 and 2023.
p-values of two-tailed tests testing the null hypothesis that the means are equal are shown in brackets.

Table 2: Baseline balance test (Section D vs Section C)

We test baseline balancing by comparing math and English placement tests, and the share of women sbetween Sections C and D. The results are shown in Table 2. These results support the random assignment of students to the two sections.

4 Empirical Specification

Our empirical specification for the effects of flipped classrooms starts with the following pooling of 2002 and 2003 data:

$$Y_{ij} = \alpha + \beta F_{ij} + \sum_{j=1}^{6} \theta_j L_j + \sum_{k=1}^{3} \gamma_k X_{ki} + \delta D_i + \phi T_i + \psi D_i T_i + \epsilon_{ij} \tag{1}$$

The dependent variable Y_{ij} is the quiz score of student i in lesson j. The dummy variable F_{ij} takes a value of one for students in flipped classrooms and zero in traditional classrooms. L_j is the lesson dummy. X_{ki} are student characteristics (i.e., Math and English

placement test scores; gender), D_i is the Section D dummy, and T_i is the year 2003 dummy. The error term is denoted as ϵ_{ij}. In this specification, the parameter of the interest β is identified off of students with and without the intervention F_{ij} in the same lesson L_j. We include additional control variables (i.e., X_{ki}, D_i, and T_i) that are not common in standard randomized controlled trials. We use cluster robust standard errors at the student level.[3]

To investigate the effect of group member characteristics in flipped classrooms, we add an interaction term with group characteristic variables, G_{ij}. When we examine the group member characteristics effects, a potential concern is the absence of students. G_{ij} is the group characteristic of student i, which is constructed from the characteristics of student i's group members who attended lesson j. As an unobserved characteristic of student i may affect the attendance of their group members as well as their quiz scores , the coefficient on G_{ij} might be estimated with bias. To mitigate such endogeneity bias, we instrument the attending group member characteristics (G_{ij}) with initially randomly assigned baseline group members' characteristics ($GAssigned_{ij}$).

For the first-stage regression, we use the following equation:

$$G_{ij} = \alpha + \mu GAssigned_{ij} + \sum_{j=1}^{6} \theta_j L_j + \sum_{k=1}^{3} \gamma_k X_{ki} + \delta D_i + \phi T_i + \psi D_i T_i + u_{ij}. \quad (2)$$

The second-stage regression is:

$$Y_{ij} = \alpha + \beta F_{ij} + \beta_{\text{Group}} G_{ij} F_{ij} + \sum_{j=1}^{6} \theta_j L_j + \sum_{k=1}^{3} \gamma_k X_{ki} + \delta D_i + \phi T_i + \psi D_i T_i + \varepsilon_{ij}. \quad (3)$$

Based on IV regression, the coefficient β_{Group} on the interaction term ($G_{ij}F_{ij}$) captures the heterogeneous effect based on group member characteristics. While β represents an average impact of the flipped classroom regardless of the group characteristics, β_{Group} captures variations in the impact due to the differences in flipped classroom's group-member

[3]Ichino et al. (2025) use 2022 data to conduct a mediation effect analysis (Imai et al., 2010) and heterogeneous treatment effects by student's attitude toward group learning, learning motivation, and self-regulated learning strategies. Furthermore, the same analysis has been conducted using both 2022 and 2023 data and qualitatively consistent findings have been found. The details are available in AppendixA1.

characteristics (i.e., maximum value of the characteristics of the group members except student i). We use cluster robust standard errors at the group level.

In our empirical specification, we include an interaction term $G_{ij}F_{ij}$ but exclude G_{ij}. This is because G_{ij} take the variations in group activities that only occur during the flipped classrooms. We estimate the model using two-stage least squares and obtain robust standard errors through a bootstrap with 1,000 replications.

5 Results and Discussion

Column (1) of Table A1 in the appendix shows the regression results of equation 1, examining the effect of the flipped classrooms on student learning. On average, across the entire sample, quiz scores in flipped classes are 0.2 standard deviations higher than in traditional classes of the same lesson. Students with higher basic math or English proficiency had statistically significantly higher quiz scores (0.1 standard deviations); no gender difference is found.

Columns (1) of Tables 3 and 4 show the OLS estimates of equation 3, which examines the heterogeneous effects of flipped classrooms due to group member characteristics. Column (2) of Tables 3 and 4 report the IV estimates of equation 3. The different panels show the results of using different measures of group member characteristics (i.e., math placement score, English placement score, attitude toward group learning, and learning motivation). The parameter of interest is the coefficient of the cross term between group member characteristics and the dummy variable for treatment (i.e., Group$\times$ Flipped). The cross terms with placement test scores and attitudes toward group learning are insignificant, but the cross term with learning motivation is statistically significant. Combined with the coefficient on the treatment (i.e., Flipped), this result suggests that the presence of group members who are highly motivated to learn in college improves their peer students'

learning outcomes in flipped classes. In particular, the effect of flipped classes is positive ($-2.229 + 0.622 \times 4 = 0.259$) when the group member's maximum score of motivation to learn is the highest. However, when the group member's maximum score of motivation to learn is not high, the effect of flipped classes is not positive ($-2.229 + 0.622 \times 3 = -0.363$). Such a pattern is not observed in the OLS estimation. The coefficient on the treatment is insignificant and that of the cross-term is significant at 10 percent. Furthermore, by comparing Columns (1) and (2) for the cross-term of learning motivation and treatment, the coefficient of the OLS estimator is smaller compared to the IV estimator. We interpret that lesson-attending group members' learning motivation has a downward bias.

6 Conclusion

With the proliferation of new technologies in education, such as video lectures and smartphone quiz apps, interest in active learning methods in higher education has grown, particularly since the COVID-19 pandemic. However, few studies have examined "what works" and "how it works" in higher education, especially outside the United States. This study aims to investigate whether the positive impact of flipped classrooms on learning outcomes in large classrooms stems from group learning effects. Specifically, we hypothesize that the characteristics of discussion group members may positively influence student learning in flipped classrooms. We seek to present a counter-argument to the typical criticism that the positive effects of flipped classrooms—beyond the video-watching component—are solely due to compulsory problem-solving exercises (such as preparing for quizzes) with support from teaching assistants and the lecturer.

Our findings indicate that the presence of highly motivated group members statistically significantly enhances students' learning outcomes in flipped classrooms, while group members' placement test scores have no notable effect. This suggests that some of the positive

effects of flipped classrooms on learning outcomes are possibly due to group work during class time. These results raise important questions regarding effectively designing group learning environments by strategically grouping students based on various non-cognitive attributes.

The limitations of the study are as follows. First, we only compare flipped classrooms and traditional lectures as a package. In a real classroom setting, it is impossible to control all conditions except the timing of the lecture (e.g., prohibiting students in traditional lectures from watching lecture videos, even if they wish to). Second, two types of history dependence could affect the results. One is the order of micro/macro and heterodox classes, and the other is the order of flipped and traditional classrooms. Third, the cluster size is not large. Finally, as with any experimental study, there is a potential threat to external validity due to the small sample size, which, although fairly representative of a private large-scale university in the country, is still limited compared to the overall university population.

Group member's characteristics based on baseline placement score	OLS (1)	IV (2)
Placement Math score		
Flipped	0.277	0.299
	(0.046)	(0.045)
Math Placement Test Score	0.112	0.112
	(0.023)	(0.017)
English Placement Test Score	0.137	0.137
	(0.024)	(0.018)
Women	-0.071	-0.071
	(0.044)	(0.036)
Group × Flipped	-0.035	-0.054
	(0.035)	(0.035)
N	3749	3749
R^2	0.063	0.064
First stage F-value		10.71
Placement English score		
Flipped	0.246	0.246
	(0.054)	(0.051)
Math Placement Test Score	0.112	0.112
	(0.023)	(0.017)
English Placement Test Score	0.136	0.136
	(0.024)	(0.018)
Women	-0.072	-0.072
	(0.044)	(0.037)
Group × Flipped	-0.002	-0.001
	(0.046)	(0.042)
N	3749	3749
R^2	0.063	0.063
First stage F-value		10.30

Robust standard errors clustered at the group level are in parentheses

All specifications include section dummy, year dummy, and the interaction term of section and year dummy.

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

Table 3: Heterogeneous effect in group member's characteristics (baseline survey questions)

Group member's characteristics based on baseline placement score	OLS (1)	IV (2)
Attitude toward group learning		
Flipped	0.619*	0.602*
	(0.297)	(0.278)
Math Placement Test Score	0.110***	0.111***
	(0.023)	(0.017)
English Placement Test Score	0.138***	0.137***
	(0.024)	(0.018)
Women	-0.071	-0.071
	(0.044)	(0.036)
Group × Flipped	-0.086	-0.081
	(0.067)	(0.063)
N	3749	3749
R^2	0.064	0.063
First stage F-value		10.59
Learning motivation		
Flipped	-0.800	-2.229**
	(0.452)	(0.727)
Math Placement Test Score	0.111***	0.110***
	(0.023)	(0.017)
English Placement Test Score	0.137***	0.137***
	(0.024)	(0.018)
Women	-0.073	-0.074*
	(0.044)	(0.036)
Group × Flipped	0.264*	0.622***
	(0.113)	(0.182)
N	3749	3749
R^2	0.065	0.066
First stage F-value		10.76

Robust standard errors clustered at the group level are in parentheses

All specifications include section dummy, year dummy and interaction term of section and year dummy.

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

Table 4: Heterogeneous effect in group member's characteristics (baseline survey questions)

References

Balaban, Rita A, Donna B Gilleskie, and Uyen 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.

Calimeris, Lauren and Katherine M Sauer (2015) "Flipping out about the flip: All hype or is there hope?" *International Review of Economics Education*, 20, 13–28.

Carrell, Scott E., Richard L. Fullerton, and James E. West (2009) "Does Your Cohort Matter? Measuring Peer Effects in College Achievement," *Journal of Labor Economics*, 27 (3), 439–464.

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

Craft, Erik and Maia Linask (2020) "Learning effects of the flipped classroom in a principles of microeconomics course," *The Journal of Economic Education*, 51 (1), 1–18.

Cunha, Flavio and James J. Heckman (2008) "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 43 (4), 738–782.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach (2006) "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *The Econometrica*, 74 (6), 1171–1230.

Dong, Xiaoyan, Yifan Liang, and Shengyu Yu (2023) "Middle-achieving students are also my peers: The impact of peer effort on academic performance," *Labour Economics*, 80, 102310.

Heckman, James J. and Tim Kautz (2012) "Hard Evidence on Soft Skills," *Labour Economics*, 19 (4), 451–464.

Heckman, James J. and Yona Rubinstein (2001) "The Importance of Noncognitive Skills: Lessons from the GED Testing Program," *American Economic Review*, 91 (2), 145–149.

Heckman, James J., Jora Stixrud, and Sergio Urzua (2006) "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24 (3), 411–482.

Hong, Sok Chul and Jungmin Lee (2017) "Who is sitting next to you? Peer effects inside the classroom," *Quantitative Economics*, 8 (2), 255–282.

Ichino, Yasukazu, Kawai Toru, and Mai Seki (2025) "Practice and Evaluation of the Flipped Classroom in Large Classes at a Higher Education Institute: A case of an Introductory Economics Course," *Japanese Journal of Educational Technology*, 49 (1), 1–18.

Imai, Kosuke, Luke Keele, and Teppei Yamamoto (2010) "Identification, inference, and sensitivity analysis for causal mediation effects," *Statistical Science*, 25 (1), 51–71.

Kimbrough, Erik O., Andrew D. McGee, and Hitoshi Shigeoka (2022) "How Do Peers Impact Learning? An Experimental Investigation of Peer-to-Peer Teaching and Ability Tracking," *Journal of Human Resources*, 57 (1), 304–339.

Kremer, Michael and Dan Levy (2008) "Peer Effects and Alcohol Use among College Students," *Journal of Economic Perspectives*, 22 (3), 189–206.

Li, Tao, Li Han, Linxiu Zhang, and Scott Rozelle (2014) "Encouraging Classroom Peer Interactions: Evidence from Chinese Migrant Schools," *Journal of Public Economics*, 111, 29–45.

Olitsky, Neal H and Sarah B Cosgrove (2016) "The better blend? Flipping the principles of microeconomics classroom," *International Review of Economics Education*, 21, 1–11.

Sacerdote, Bruce (2001) "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *The Quarterly Journal of Economics*, 116 (2), 681–704.

Swoboda, Aaron and Lauren Feiler (2016) "Measuring the effect of blended learning: Evidence from a selective liberal arts college," *American Economic Review*, 106 (5), 368–372.

Wang, Chunchao, Aiping Xiao, and Yu Zhou (2024) "Teamwork and Human Capital Development," *Journal of Human Resources*, 59 (5), 1425–1457.

Wozny, Nathan, Cary Balser, and Drew Ives (2018) "Evaluating the flipped classroom: A randomized controlled trial," *The Journal of Economic Education*, 49 (2), 115–129.

Zimmerman, David J. (2003) "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *The Review of Economics and Statistics*, 85 (1), 9–23.

A Medication Analysis

Column (1) of Table A1 shows the effects of flipped classrooms on student learning. Further, we replicate a mediation analysis with data in 2022 and 2023. The results are shown in Table A2. The mediation analysis is based on the results of Columns (2) and (3) of Table A1.

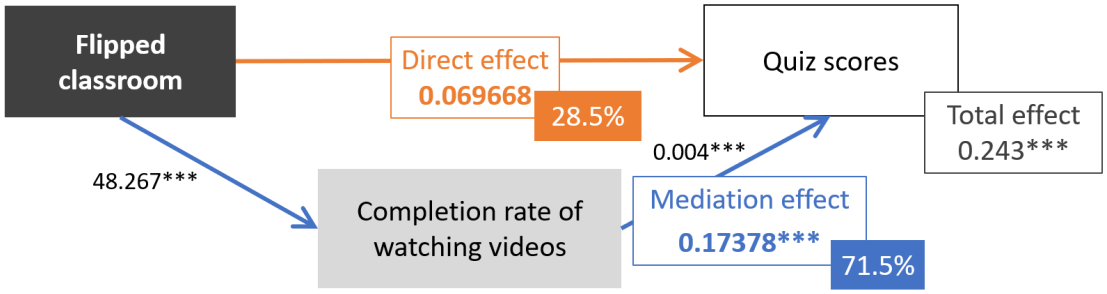

Figure A1: Mediation analysis via video completion rate

		Regression for mediation analysis	
	Quiz	Watching videos	Quiz
	(1)	(2)	(3)
Flipped	0.243***	48.267***	0.070
	(0.032)	(1.616)	(0.043)
Watching video Completion Rate			0.004***
			(0.000)
Math Placement Test Score	0.112***	-0.805	0.115***
	(0.024)	(0.984)	(0.022)
English Placement Test Score	0.136***	3.484***	0.124***
	(0.025)	(0.987)	(0.023)
Women	-0.072	3.250	-0.083*
	(0.046)	(2.040)	(0.043)
N	3749	3749	3749
R^2	0.063	0.300	0.081

Column (1) shows effect of flipped classrooms on quiz results.

Robust standard errors clustered at the student level are in parentheses.

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

Table A1: Effects of flipped classrooms and regression results for mediation analysis

	Quiz
Mediation effect	0.174^{***}
	(0.021)
Direct effect	0.070
	(0.043)
Total effect	0.243^{***}
	(0.035)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2: Result for the mediation analysis