国立大学法人 一橋大学 森有礼高等教育国際流動化機構

Working Paper Series
Mori Arinori Institute for Higher Education and Global Mobility

No.WP2021-03

The Effects of Large-Scale Online Classes on Students' Course Evaluations: Evidence from a Japanese University under the COVID-19 Pandemic

Ryohei Kashima　　Yohei Yamamoto

2021年8月

HITOTSUBASHI UNIVERSITY

The Effects of Large-Scale Online Classes on Students' Course Evaluations: Evidence from a Japanese University under the COVID-19 Pandemic*

Ryohei Kashima†
Hitotsubashi University

Yohei Yamamoto‡
Hitotsubashi University

August 16, 2021

Abstract

We investigate the effects of an introduction of large-scale online classes under the COVID-19 pandemic on students' course evaluations using questionnaire data in a Japanese university from 2018-2020. We employ difference-in-differences analysis with means-as-outcomes using means at the student and instructor levels, respectively. Students' course evaluations improved for most question items when the raw questionnaire data are used. However, the improvements are weak or none if the instructor-level means are used. These results suggest that improvements in course evaluations are less attributed to teaching quality but more to patterns of students' class-taking.

JEL Classification: A22, I21

Keywords: COVID-19; Teaching Quality; Means-as-Outcomes; Difference-in-Differences

---

*We appreciate the useful comments from participants of the HIAS lunch seminar. All errors are our own.

†Graduate School of Economics and Mori Arinori Institute for Higher Education and Global Mobility, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo, 186-8601 (rkashi1228@gmail.com).

‡Graduate School of Economics and Mori Arinori Institute for Higher Education and Global Mobility, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo, 186-8601 (yohei.yamamoto@r.hit-u.ac.jp).

1 Introduction

Universities across nations introduced large-scale online classes in the year of the COVID-19 pandemic. There is a strong concern as to how they have affected the students' course evaluations and, ultimately, the quality of teaching.[1] This study approaches such questions using data of regularly conducted student course evaluations of an anonymous Japanese university from 2018-2020. Using regular questionnaire data is advantageous in identifying causal effects, as the data are available from before the COVID-19. This helps overcome the non-presence of a valid "control group" even when the entire university began implementing online classes, because the data before COVID-19 can be regarded as a control group. As such, we employ difference-in-differences (DID) analysis by taking the change from 2019-2020 as the treatment group and the most recent change from 2018-2019 as a control group.

Our goal is to disentangle multiple effects that occurred under the COVID-19 pandemic on students' course evaluations. In particular, we focus on the following two major effects. First, teaching online can affect course evaluations through the quality of teaching. Some reasons for this include the following widely concerned facts: online classes may be considered to discourage students in having discussions with the instructor and classmates, undermining the strictness of examinations and evaluation fairness, making working in groups more difficult, and creating a heavier homework burden.[2] Second, large-scale online classes under the

[1] There is rapidly increasing literature that is concerned with the negative impacts of the COVID-19 countermeasures on university education. Orlov et al. (2021) showed that test scores declined after the COVID-19 outbreak, and Aucejo et al. (2020) have raised the concern that many students had delayed graduation and lost job opportunities. Agostinelli et al. (2020) point out that a temporary loss of educational opportunities could have a long-run negative effect. Andrew et al. (2020) found that students spend a shorter amount of time studying at home under COVID-19, and this is more pronounced for students from low-income households. Engzell et al. (2020) warn that even students in countries with highly developed information infrastructure lose educational opportunities. However, Bacher-Hicks et al. (2020) evaluated rapid accumulations in online educational resources under the COVID-19 pandemic, and Ikeda and Yamaguchi (2021) propose using online learning services to compensate for the loss of educational opportunities for Japanese high school students during the school closures.

[2] More fundamentally, the effects of online teaching were discussed even before the COVID-19 outbreak. Coates et al. (2004) and Figlio et al. (2013) showed that when the same instructor gives the same class face to face and online, the students who took the online class had lower test scores. Alpert et al. (2016) conducted the same class fully face to face, fully online, and a mixture of the two and found that students' performance in the online classes was significantly lower than that in the face-to-face classes, and there was no difference between the mixed and face-to-face classes. Using non-experimental data but employing the instrumental variable method, Bettinger et al. (2017) showed that online classes lower exam scores; hence, they may lead to absence or withdrawal from school. Although the average effect is negative, Bettinger et al. (2017) found that the negative effect of online classes disappears for students with a high prior college GPA, and Figlio et al. (2013) found similar results for higher high school GPA students. Furthermore, Cacault et al. (2021) showed that taking classes through a livestreaming format had a positive effect on students with a

COVID-19 pandemic changed patterns of students' class-taking because physical restrictions of students (such as time) and universities (such as classroom capacities) were lessened.

We use data of non-mandatory course evaluations. In addition to the DID analysis using the raw questionnaire data, we estimate the model using two different types of means-as-outcomes to consider changes in patterns of students' class-taking that might have occurred in 2020. First, we use the student-level means as they adjust for the tendency that students who tend to give higher evaluations take more courses and thus give more responses. Second, we use the instructor-level means as they adjust for the tendency that students take more courses and thus give more responses for highly rated instructors. These analyses are useful to control for patterns of students' class-taking when teaching quality may also change.

Our findings are summarized as follows. First, results using the raw questionnaire data and using the student-level means show improvements in students' course evaluations in all question items but one. However, the improvements become weak or none if we use the instructor-level means. These results suggest that the overall improvements of the raw questionnaire mean are less attributed to teaching quality but more to patterns of students' class-taking. In the remainder of this paper, we describe the data in section 2, explain our empirical strategy in section 3, report our results in section 4, and conclude in section 5.

2 Data

We use repeated cross-sections of students' course evaluations from an anonymous Japanese university for the spring and summer quarters of the academic years 2018-2020. The survey is a web-based questionnaire that students can answer for all the courses they take toward the course's end. The response rates were 31.1% in 2018, 32.0% in 2019, and 30.1% in 2020, which concern us regarding selection bias in estimating the population mean. However, since we focus on the difference between two years, the selection bias is expected to be offset if the sources are unchanged, and the difference estimator will not be much distorted.[3]

Table 1 shows the number of responses to the survey and summary statistics (mean,

high GPA in high school. These results demonstrate the existence of heterogeneity in the online class effect. Other studies, such as those by Chen and Lin (2016), Heissel (2016), and Escueta et al. (2017), found that online classes are not perfect substitutes for face-to-face classes.

[3] As an analogy, Perron and Yamamoto (2015) discuss the same issue in the context of a structural change model. Even if endogeneity is present, as long as the magnitudes and the dates of structural changes are taken into account, the simple ordinary least squares (OLS)-based estimator is consistent under certain assumptions. Moreover, the OLS-based estimator is more precise than the instrumental variables (IV)-based estimator in general. The tests for structural changes based on the OLS-based estimator have a higher power than the tests based on the IV-based estimator.

standard deviation, skewness, and kurtosis) for each year by grade. The total number of responses for the 3 years is 26,583, with a tendency of fewer responses from students in upper grades. We used seven question items commonly asked in the 3 years: frequency of student attendance ("Attendance"), time spent studying the subject outside the class ("Study hours outside the class"), how clear the course objectives are ("Clearness of objectives"), transparency of the grading criteria ("Clearness of evaluation"), how understandable the instructor's explanations are ("Clearness of explanation"), how enthusiastic the instructor is ("Enthusiasm of the instructor"), and how useful the student feels the course is ("Significance of the course"). For each question, the answer is selected from 1-5, and the higher the score, the more positive the evaluation. [4] The mean of "Attendance" decreases, and its standard deviation increases as students go to the upper grades. The mean and standard deviation for "Study hours outside the class" decreases as students go to the upper grades. For the other questions, the mean slightly increases, and the standard deviation decreases as the students go to the upper grades. For "Attendance", the number of students who rated 5 increased, and thus the skewness and the kurtosis are severe in 2020, although the sample size is large enough to disregard these effects in the following statistical analysis.

3 Empirical strategy

3.1 Difference-in-differences

To assess the effect of large-scale online classes, it is useful to compare the change in the averages of responses from 2019-2020 with that when they were not introduced but the same otherwise. The former works as the treatment group, and the latter works as the control group. Our particular choice of the latter is the sample that is as recent as possible, that is, the change from 2018-2019. Hence, we estimate the following quantity as the DID effect:

$$\{E(Y_i|2020, Z_i) - E(Y_i|2019, Z_i)\} - \{E(Y_i|2019, Z_i) - E(Y_i|2018, Z_i)\}. \tag{1}$$

To this end, we estimate the following regression model

$$Y_i = \beta_0 + \beta_1 D20_i + \beta_2 D18_i + Z_i'\gamma + e_i, \tag{2}$$

for $i = 1, ..., N$, where Y_i is the raw questionnaire response, $D18_i$ is a dummy variable that takes 1 if the response is in 2018 and 0 otherwise, $D20_i$ is a dummy variable that takes the

[4] For "Study hours outside of class", since there are differences in the classification of response items between school year 2018 and 2019-2020, we converted the items into a three-level evaluation classification that allows us to integrate the responses from the 3 years.

value of 1 if the response is in 2020 and 0 otherwise, Z_i is a vector of the control variables, and e_i is the error term with mean zero. For the control variables Z_i, we include dummy variables for the major of student, grade of the student, courses, and instructors.

Since we used 2019 as the reference case in (2), the conditional expectation of Y_i in terms of (2) in each year is described as follows:

$$E(Y_i|2018, Z_i) = \beta_0 + \beta_2, \tag{3}$$

$$E(Y_i|2019, Z_i) = \beta_0, \tag{4}$$

$$E(Y_i|2020, Z_i) = \beta_0 + \beta_1. \tag{5}$$

Hence, the causal effect (1) can be estimated as

$$\begin{aligned} & \{E(Y_i|2020, Z_i) - E(Y_i|2019, Z_i)\} - \{E(Y_i|2019, Z_i) - E(Y_i|2018, Z_i)\}, \\ = \ & \{(\beta_0 + \beta_1) - \beta_0\} - \{\beta_0 - (\beta_0 + \beta_{,2})\}, \\ = \ & \beta_1 + \beta_2. \end{aligned}$$

Figure 1 illustrates how these quantities correspond to the DID model. If the change from 2018-2019 occurred in 2020, the causal effect was captured by $\beta_1 + \beta_2$. If not, we may regard β_1 as the causal effect, assuming that no trend exists. For the former, we can test using an F test for the null hypothesis of $H_0 : \beta_1 + \beta_2 = 0$ and the alternative hypotheses of $H_1 : \beta_1 + \beta_2 \neq 0$. For the latter, we can test for $H_0 : \beta_1 = 0$ and $H_1 : \beta_1 \neq 0$ by a t test.

3.2 Difference-in-differences with means-as-outcomes

Our response data have multiple attributes at higher levels: student, instructor, course, and year. Besides estimating the DID model using the raw questionnaire data, we focus on some categories by constructing means-as-outcomes prior to estimating the regression model (2).[5] In particular, we use two types of means as follows. First, we take the means of the responses of each student prior to estimating the DID model. We call this the student-level means. Estimation using these can adjust the tendency that students who give higher evaluations take more courses and thus give more responses in the COVID-19 year. Second, we take the means of responses that each instructor receives before estimating the DID model. We call this the instructor-level means. Estimation using these can adjust the tendency that students take more courses and thus give more responses toward highly rated instructors in

[5] See Bryk and Raudenbush (2002).

the COVID-19 year. If this is the case, the mean of the raw questionnaire data may increase, even though the teaching quality remains the same, while analysis using the instructor-level means keeps the mean unchanged.

Figure 2 illustrates our framework in a very simple example of changes in patterns of students' class-taking. This is a case in which the students take more courses and thus give more responses to highly rated instructors. Suppose that in 2019, a student gives a 4.0 to the course of instructor A, and another student gives a 2.0 to the course of instructor B. Then, the average score using the raw questionnaire data is 3.0 in 2019. Importantly, the average score is 3.0, even when the student-level and instructor-level means are used. Suppose now that in 2020, patterns of students' class-taking change, and the number of students enrolled in the courses of instructor A increase. Interestingly, even though the quality of courses remains the same, the average score in 2020 increases to 3.3 if the raw questionnaire data are used. When the student-level means are used, the average is even higher (3.5). Interestingly, the average score of the instructor-level means remains at 3.0.[6] Therefore, the student-level and instructor-level means are useful to consider such a tendency.

When we use the means-as-outcomes in (2), the control variables can differ. When we use the student-level means, the dummy variables for the student's major and the grade are included. When we use the instructor-level means, the dummy variables for the instructor's department are used, although these control variables do not affect our main results.

4 Results

Table 2 shows the results of (2) using the raw questionnaire data. The coefficient β_1, that is, (5) minus (4), represents the change in the conditional mean of the responses from 2019-2020. The estimates are positive and highly significant for all question items and indicate that the students' course evaluations improved from 2019-2020. To obtain a more appropriate causal interpretation, we investigated the DID effects $\beta_1 + \beta_2$. The estimates are again positive and significant at the 1% level for all question items, but "Clearness of evaluation" is negative but insignificant. This reflects students' anxiety about online evaluation methods compared to face-to-face proctored exams. "Attendance" shows the largest increase, which may induce a change in patterns of class-taking. Finally, the DID effects are smaller than β_1 for all the question items. This suggests an overall tendency of improvement from 2018-2019. It is important to ask whether this continues in 2020 to set an appropriate counterfactual in the

[6] We can also construct an example in which students who tend to give higher evaluations take more courses, in which the student-level mean does not change.

DID analysis. Another choice is to assume no change as a counterfactual and to interpret β_1 as a causal effect. In either case, the final conclusions are qualitatively the same.

Table 3 shows the estimation results of (2) using the student-level means. The DID effects are larger compared with those in Table 2 for many question items, such as "Attendance", "Study hours outside the class", and "Significance of the course". These suggest the tendency that students who take less courses have larger improvements in their evaluations in the COVID-19 year. Moreover, the DID effects are positive and significant at the 1% level for all question items but "Clearness of evaluation". In this sense, we can say that the results are overall similar to the previous analysis using the raw questionnaire data in Table 2.

The results using the instructor-level means are presented in Table 4. When we look at the changes from 2019-2020 by β_1, "Attendance", "Study hours outside the class", "Clearness of objectives", and "Significance of the course" are positive and significant at the 1% level; however, the values are much smaller than those using the raw questionnaire data and using the student-level means. "Enthusiasm of the instructor" and "Clearness of explanation" are positive, but the former is significant only at the 10% level, and the latter is insignificant. "Clearness of evaluation" turned negative, although it is insignificant. Interestingly, if we look at the DID effects ($\beta_1 + \beta_2$), all question items but "Attendance" are close to 0 and insignificant, and "Clearness of evaluation" and "Enthusiasm of the instructor" are negative.

The last results go a long way toward our goal. If online classes improve teaching quality, positive and significant coefficients would be observed in the instructor-level analysis too. In particular, for question items such as "Clearness of objectives", "Clearness of explanations", "Enthusiasm of the instructor", and "Significance of the course" are closely related to it. Hence, these results suggest that improvements in course evaluations are less attributed to teaching quality and more to patterns of students' class-taking.

5 Conclusion

Our results do not support the view that online classes improved teaching quality and thus students' course evaluations under the COVID-19 pandemic. Rather, the results are more consistent with existing studies that are conservative about the effect of online teaching (Cacault et al. 2021; Bettinger et al. 2017; Figlio et al. 2013). However, this does not imply that online teaching itself never improves students' course evaluations and, ultimately, teaching quality. We are sure that they do if online methods are flexibly and appropriately used in the entire education system, hopefully even in the post-COVID-19 regime.

References

Agostinelli, F., M. Doepke, G. Sorrenti, and F. Zilibotti (2020) "When the great equalizer shuts down: schools, peers, and parents in pandemic times," Working Paper 28264, National Bureau of Economic Research.

Alpert, W. T., K. A. Couch, and O. R. Harmon (2016) "A randomized assessment of online learning," *American Economic Review* 106(5), 378–382.

Andrew, A., S. Cattan, M. Costa Dias, C. Farquharson, L. Kraftman, S. Krutikova, A. Phimister, and A. Sevilla (2020) "Inequalities in children's experiences of home learning during the covid-19 lockdown in England," *Fiscal Studies* 41(3), 653–683.

Aucejo, E. M., J. F. French, M. P. U. Araya, and B. Zafar (2020) "The impact of covid-19 on student experiences and expectations: Evidence from a survey," Working Paper 27392, National Bureau of Economic Research.

Bacher-Hicks, A., J. Goodman, and C. Mulhern (2020) "Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time," Working Paper 27555, National Bureau of Economic Research.

Bettinger, E.P., L. Fox, S. Loeb, and E.S. Taylor (2017) "Virtual classrooms: How online college courses affect student success," *American Economic Review* 107 (9), 2855–2875.

Bryk, A.S. and S.W. Raudenbush (2002) "Hierarchical linear models: Applications and data analysis methods," Sage Publication, Inc., Thousand Oaks.

Cacault, M.P., C. Hidebrand, J. Laurent-Lucchetti, M. Pellizzari (2021) "Distance learning in higher education: Evidence from a randomized experiment," forthcoming in *Journal of the European Economic Association*.

Chen, J., and T.-F. Lin (2016) ". The benefit of providing face-to-face lectures in online learning microeconomics courses: Evidence from a regression discontinuity design experiment," *Economics Bulletin* 36(4), 2094–2116.

Coates, D., B. R. Humphreys, J. Kane, and M. A. Vachris (2004) ""No significant distance" between face-to-face and online instruction: evidence from principles of economics," *Economics of Education Review* 23(5), 533–546.

Engzell, P., A. Frey, and M. D. Verhagen (2020) "Learning inequality during the covid-19 pandemic," SocArXiv Papers. (DOI: 10.31235/osf.io/ve4z7)

Escueta, M., V. Quan, A. J. Nickow, and P. Oreopoulos (2017) "Education technology: An evidence-based review," Working Paper 23744, National Bureau of Economic Research.

Figlio, D., M. Rush, and L. Yin (2013) "Is it live or is it internet? Experimental estimates of the effects of online instruction on student learning," *Journal of Labor Economics* 31(4), 763–784.

Heissel, J. (2016) "The relative benefits of live versus online delivery: Evidence from virtual algebra I in North Carolina," *Economics of Education Review* 53, 99–115.

Ikeda, M. and S. Yamaguchi (2021) "Online learning during school closure due to COVID-19," *The Japanese Economic Review* 72, 471-507.

Orlov, G., D. McKee, J. Berry, A. Boyle, T. DiCiccio, T. Ransom, A. Rees-Jones, and J. Stoye (2021) "Learning during the covid-19 pandemic: It is not who you teach, but how you teach," *Economics Letters* 202, 109812.

Perron, P., and Y. Yamamoto (2015) "Using OLS to estimate and test for structural changes in models with endogenous regressors", *Journal of Applied Econometrics* 30, 119-144, 2015.

Table 1. Summary statistics

(1) From 2018 to 2020

	Number of responses	Attendance	Study hours outside the class	Clearness of objectives	Clearness of evaluation	Clearness of explanation	Enthusiasm of the instructor	Significance of the course
mean								
1st	12107	4.77	1.70	3.89	3.84	3.88	4.14	4.00
2nd	7812	4.64	1.55	4.02	3.99	4.05	4.25	4.09
3rd	4972	4.55	1.41	3.98	3.93	4.04	4.25	4.07
4th	1692	4.32	1.47	4.06	4.04	4.15	4.33	4.19
standard deviation								
1st	12107	0.65	0.70	0.88	0.93	0.99	0.83	0.92
2nd	7812	0.80	0.67	0.83,	0.89	0.92	0.78	0.89
3rd	4972	0.89	0.60	0.84	0.91	0.93	0.79	0.90
4th	1692	1.00	0.62	0.82	0.86	0.91	0.77	0.86
skewness								
1st	12107	-3.55	0.48	-0.92	-0.75	-0.89	-1.09	-1.01
2nd	7812	-2.68	0.83	-1.00	-0.91	-1.08	-1.21	-1.13
3rd	4972	-2.34	1.16	-1.09	-1.00	-1.08	-1.24	-1.14
4th	1692	-1.59	0.98	-1.07	-1.01	-1.36	-1.36	-1.27
kurtosis								
1st	12107	16.97	2.13	4.09	3.41	3.54	4.66	4.10
2nd	7812	10.42	2.54	4.54	3.87	3.87	5.25	4.51
3rd	4972	8.36	3.30	4.87	4.14	4.14	5.35	4.59
4th	1692	5.03	2.91	4.76	4.36	4.36	5.64	4.97

(2) 2020

	Number of responses	Attendance	Study hours outside the class	Clearness of objectives	Clearness of evaluation	Clearness of explanation	Enthusiasm of the instructor	Significance of the course
mean								
1st	3818	4.92	1.95	4.06	3.86	3.99	4.22	4.17
2nd	2830	4.90	1.91	4.21	4.07	4.17	4.33	4.25
3rd	1805	4.84	1.76	4.12	3.97	4.15	4.30	4.20
4th	522	4.75	1.73	4.20	4.06	4.20	4.36	4.27
standard deviation								
1st	3818	0.39	0.67	0.77	0.91	0.92	0.79	0.81
2nd	2830	0.44	0.67	0.74	0.89	0.87	0.76	0.80
3rd	1805	0.58	0.66	0.80	0.93	0.90	0.79	0.84
4th	522	0.70	0.65	0.74	0.89	0.93	0.78	0.86
skewness								
1st	3818	-6.23	0.06	-0.96	-0.77	-0.98	-1.14	-1.10
2nd	2830	-5.67	0.11	-1.15	-1.02	-1.30	-1.40	-1.32
3rd	1805	-4.67	0.32	-1.18	-1.03	-1.19	-1.35	-1.23
4th	522	-3.53	0.34	-1.27	-1.22	-1.55	-1.61	-1.55
kurtosis								
1st	3818	49.25	2.26	4.72	3.51	3.94	5.00	4.70
2nd	2830	40.83	2.20	5.63	4.13	5.16	6.11	5.61
3rd	1805	27.24	2.22	5.27	4.06	4.38	5.65	4.92
4th	522	16.57	2.25	6.32	4.90	5.63	6.76	-1.55

Table 1. Summary statistics (continued)

(3) 2019

	Number of responses	Attendance	Study hours outside the class	Clearness of objectives	Clearness of evaluation	Clearness of explanation	Enthusiasm of the instructor	Significance of the course
mean								
1st	3954	4.70	1.70	3.85	3.82	3.80	4.10	3.93
2nd	2798	4.55	1.48	3.98	3.96	4.02	4.24	4.04
3rd	1770	4.47	1.36	4.00	3.97	4.03	4.28	4.07
4th	599	4.17	1.38	4.01	4.02	4.10	4.33	4.15
standard deviation								
1st	3954	0.73	0.70	0.91	0.96	1.03	0.84	0.94
2nd	2798	0.88	0.65	0.83	0.88	0.93	0.76	0.89
3rd	1770	0.92	0.57	0.81	0.90	0.92	0.75	0.86
4th	599	1.02	0.59	0.87	0.84	0.91	0.76	0.87
skewness								
1st	3954	-3.03	0.49	-0.89	-0.72	-0.83	-1.00	-0.98
2nd	2798	-2.34	1.02	-0.93	-0.87	-1.00	-1.07	-1.06
3rd	1770	-2.03	1.33	-1.05	-1.05	-1.05	-1.18	-1.05
4th	599	-1.24	1.29	-0.98	-0.82	-1.18	-1.12	-1.20
kurtosis								
1st	3954	13.08	2.12	3.90	3.25	3.31	4.35	3.94
2nd	2798	8.38	2.88	4.35	3.85	3.96	4.84	4.36
3rd	1770	6.88	3.78	5.00	4.39	4.18	5.38	4.53
4th	599	4.12	3.62	4.23	3.87	4.51	4.41	4.80

(4) 2018

	Number of responses	Attendance	Studyhours outside the class	Clearnessof objectives	Clearnessof evaluation	Clearnessof explanation	Enthusiasmof the instructor	Significanceof thecourse
mean								
1st	4335	4.67	1.42	3.75	3.85	3.82	4.10	3.89
2nd	2184	4.53	1.33	3.91	3.95	3.98	4.21	4.01
3rd	1397	4.40	1.19	3.85	3.86	3.96	4.18	3.97
4th	571	4.04	1.27	3.98	4.03	4.16	4.30	4.15
standard deviation								
1st	4335	0.77	0.61	0.94	0.93	1.01	0.87	0.98
2nd	2184	0.88	0.55	0.87	0.91	0.94	0.82	0.94
3rd	1397	0.99	0.42	0.88	0.91	0.96	0.82	0.96
4th	571	1.11	0.49	0.84	0.85	0.89	0.77	0.85
skewness								
1st	4335	-2.87	1.15	-0.79	-0.76	-0.83	-1.09	-0.88
2nd	2184	-2.18	1.43	-0.93	-0.87	-1.00	-1.19	-1.03
3rd	1397	-1.89	2.02	-1.06	-0.93	-1.03	-1.19	-1.11
4th	571	-1.11	1.58	-0.94	-0.95	-1.38	-1.34	-1.07
kurtosis								
1st	4335	11.67	3.25	3.60	3.47	3.38	4.58	3.63
2nd	2184	7.64	4.08	4.15	3.74	3.93	5.04	4.05
3rd	1397	6.18	6.23	4.54	4.03	4.06	5.05	4.28
4th	571	3.50	4.56	4.27	4.18	5.32	5.75	4.20

Table 2. Difference-in-differences estimation using the raw questionnarie data

	Attendance	Study hours outside the class	Clearness of objectives	Clearness of evaluation	Clearness of explanation	Enthusiasm of the instructor	Significance of the course
β_1	0.329***	0.299***	0.245***	0.046***	0.207***	0.100***	0.234***
	(0.014)	(0.011)	(0.016)	(0.017)	(0.017)	(0.015)	(0.017)
β_2	-0.072***	-0.249***	-0.137***	-0.063***	-0.053***	-0.028*	-0.106***
	(0.016)	(0.012)	(0.017)	(0.019)	(0.019)	(0.016)	(0.018)
$\beta_1+\beta_2$	0.257***	0.050***	0.108***	-0.017	0.154***	0.072***	0.128***
$F\ stat$	105.25	6.728	14.957	0.326	26.312	7.561	18.864
$p-value$	0.000	0.009	0.000	0.568	0.000	0.006	0.000
$adj.R^2$	0.146	0.315	0.118	0.097	0.187	0.128	0.121
N	26583	26583	26583	26583	26583	26583	26583

Note: Standard errors are shown in parentheses. ***, **, and * indicate significance at the 1%, the 5%, and the 10% levels, respectively. "$F\ stat$" shows the F statistic for $H_0 : \beta_1 + \beta_2 = 0$ against $H_1 : \beta_1 + \beta_2 \neq 0$. The regression includes dummy variables for the major of student, grade of the student, courses, and instructors, as control variables but their coefficients are suppressed.

Table 3. Difference-in-differences estimation using the student-level means

	Attendance	Study hours outside the class	Clearness of objectives	Clearness of evaluation	Clearness of explanation	Enthusiasm of the instructor	Significance of the course
β_1	0.377***	0.357***	0.195***	0.054***	0.171***	0.102***	0.215***
	(0.019)	(0.015)	(0.020)	(0.021)	(0.021)	(0.018)	(0.019)
β_2	-0.045***	-0.184***	-0.089***	-0.017***	-0.014***	-0.017*	-0.031***
	(0.018)	(0.014)	(0.019)	(0.020)	(0.020)	(0.018)	(0.019)
$\beta_1+\beta_2$	0.332***	0.173***	0.106***	0.037	0.157***	0.085***	0.184***
$Fstat$	114.29	50.522	10.378	1.151	27.692	7.520	29.525
$p-value$	0.000	0.000	0.001	0.283	0.000	0.006	0.000
$adj.R^2$	0.122	0.203	0.044	0.016	0.039	0.029	0.036
N	7354	7354	7354	7354	7354	7354	7354

Note: Same as in Table 2 except that the control variables include the dummy variables for the student's major and grade.

Table 4. Difference-in-differences estimation using the instructor-level means

	Attendance	Study hours outside the class	Clearness of objectives	Clearness of evaluation	Clearness of explanation	Enthusiasm of the instructor	Significance of the course
β_1	0.270***	0.261***	0.137***	-0.001	0.088*	0.037	0.127***
	(0.032)	(0.034)	(0.034)	(0.038)	(0.043)	(0.037)	(0.038)
β_2	-0.062***	-0.196***	-0.114***	-0.056	-0.050	-0.042	-0.050
	(0.032)	(0.032)	(0.033)	(0.036)	(0.042)	(0.035)	(0.037)
$\beta_1+\beta_2$	0.208***	0.065	0.023	-0.057	0.038	-0.005	0.077
$Fstat$	14.984	1.307	0.162	0.787	0.271	0.007	1.378
$p-value$	0.000	0.253	0.688	0.375	0.603	0.931	0.241
$adj.R^2$	0.167	0.264	0.072	-0.001	0.033	0.020	0.054
N	896	896	896	896	896	896	896

Notes: Same as in Table 2 except that the control variables include the dummy variables for the instructor's department.

Figure 1. Difference-in-differences analysis

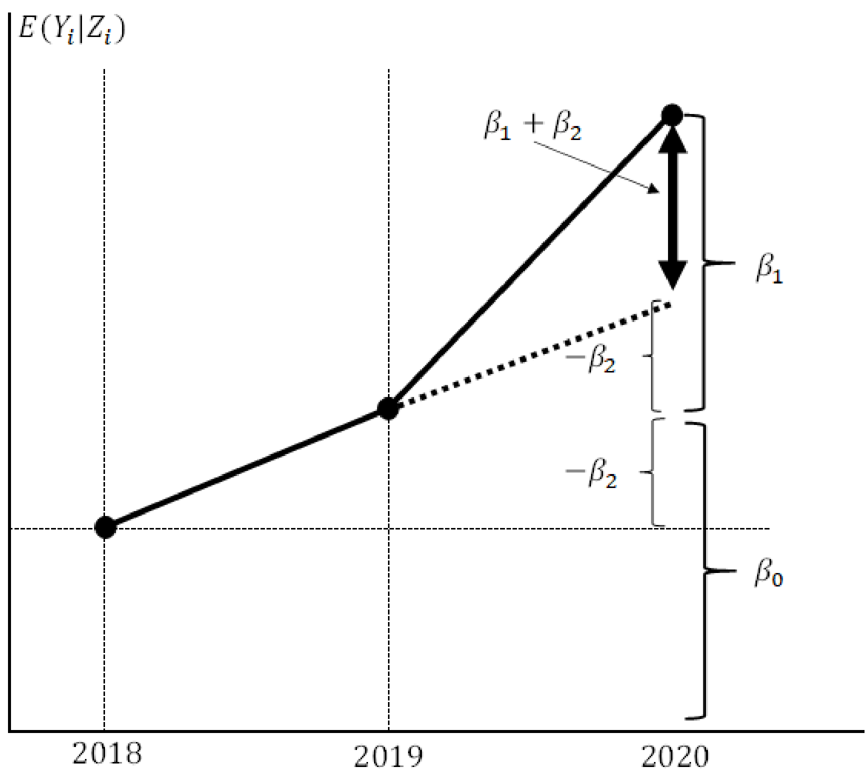

Figure 2. Effects of students' class-taking on average scores

2019

instructor	student
A	4
B	2

Average scores

questionnaire data	3.0
student-level	3.0
instructor-level	3.0

2020

Questionnaire-level

instructor	student	
A	4	4
B	2	

Student-level

instructor	student
A	4
B	3

Instructor-level

instructor	student
A	4
B	2

Average scores

questionnaire data	3.3
student-level	3.5
instructor-level	3.0