Learning during the COVID-19 pandemic: It is not who you teach, but how you teach

Authors: George Orlov[1]*, Douglas McKee[1]*, James Berry[2], Austin Boyle[3], Thomas J. DiCiccio[4], Tyler Ransom[5], Alex Rees-Jones[6], Jörg Stoye[1]

Affiliations:

[1]Cornell University, Department of Economics

[2]University of Delaware, Department of Economics

[3]Pennsylvania State University, Department of Economics

[4]Cornell University, Department of Social Statistics

[5]University of Oklahoma, Department of Economics

[6]University of Pennsylvania, The Wharton School, Business Economics and Public Policy Department

*Correspondence to: george.orlov@cornell.edu, douglas.mckee@cornell.edu

Abstract: We use standardized end-of-course knowledge assessments to examine student learning during the disruptions induced by the COVID-19 pandemic. Examining seven economics courses taught at four US R1 institutions, we find that students performed substantially worse, on average, in Spring 2020 when compared to Spring or Fall 2019. We find no evidence that the effect was driven by specific demographic groups. However, our results suggest that teaching methods that encourage active engagement, such as the use of small group activities and projects, played an important role in mitigating this negative effect. Our results point to methods for more effective online teaching as the pandemic continues.

Acknowledgments: George Orlov and Douglas Mckee conceptualized the project, compiled the data, performed the analysis, and wrote the initial draft of the paper. Berry, Boyle, DiCiccio, Ransom, Rees-Jones, and Stoye participated in vigorous discussions of all aspects of the paper including methodology, analysis, and writing. All authors except Orlov contributed data from classes that they taught in Spring 2019, Fall 2019, and/or Spring 2020. The project did not rely on any formal research funding. The authors would like to thank Carolyn Aslan and Amy Cardace from the Cornell Center for Teaching Innovation for their support during the IRB process as well as Peter LePage who encouraged our work as the head of Cornell's Active Learning Initiative.

When the COVID-19 pandemic arrived in the United States in the spring of 2020, most colleges and universities switched from in-person teaching to remote instruction. As the pandemic continues to unfold, even those institutions that brought students back to campus in the Fall 2020 term have had to offer substantial numbers of courses online. For many institutions, this transition to online learning was conducted on short notice, with little planning or prior experience to guide the transitions. For educational institutions to be successful in providing students with the best possible learning experience in this new environment, it is essential to understand which aspects of pedagogy proved to be most effective and whether specific groups of students were more vulnerable in the forced switch to remote instruction, so that they can be provided with additional support.

Investigating how different aspects of teaching affect the learning of different types of students is often challenging. Typically, our best measure of learning in a course is the final exam, and these exams can differ in difficulty or not evaluate the same course learning goals from semester to semester. In the pandemic, these challenges are further complicated by changes in the way final exams are often administered (e.g., going from a closed book proctored exam taken on campus to an open book unproctored exam taken online in a student's home). We circumvent this issue by analyzing data from seven intermediate-level economics courses in which student learning was measured using standard multiple-choice assessments developed at Cornell University as a part of the Active Learning Initiative (*1)*, following the procedure outlined in (*2*): the Intermediate Economics Skills Assessment – Microeconomics (IESA-Micro, 31 questions), the Economic Statistics Skills Assessment (ESSA, 20 questions), the Applied Econometrics Skills Assessment (AESA, 24 questions), and the Theory-based Econometrics Skills Assessment (TESA, 21 questions). Each of the assessment questions are mapped to explicit course learning goals, and assessments were administered as low-stakes tests just prior to or just after the final class meeting of each semester.

In this paper, we compare student performance on standard assessments in Spring 2020 to student performance in the same courses in either Fall or Spring 2019 to estimate the impact of the emergency switch to remote instruction induced by the COVID-19 pandemic. Using these data, we address three questions: First, we examine how end-of-semester knowledge was influenced by the measures taken in Spring 2020. Second, we assess whether certain groups of students were more affected by the pandemic.[1] And third, we look at whether the use of specific teaching methods resulted in a more successful transition to remote teaching.

Our data were collected during the Spring 2019, Fall 2019, and Spring 2020 semesters at four R1 PhD-granting institutions. Student data include the performance on the multiple-choice assessments and responses to a demographic questionnaire. At the end of the Spring 2020 semester, instructors of the seven courses filled out a survey regarding their teaching practices before and during the pandemic and the extent of material coverage during the pandemic semester. All but one of the seven courses were taught by the same instructor in the pre-pandemic and pandemic semesters. Since each of the assessment questions is mapped to one or more course-specific learning goals, we were able to calculate a separate subscore for the material that was taught remotely during the latter portion of the semester. Our analysis sample pools the students who completed the study courses with two sets of restrictions imposed: First, students must have answered survey questions on gender, ethnicity, parental education, and non-

[1] This question is partially motivated by prior findings that African American students and those with lower grade point averages perform worse in online classes than in-person classes *(3).*

native English speaker status. Response rates varied somewhat across courses, but based on administrative data, it does not look like changes in rates across semesters in the same courses were correlated with student characteristics such as GPA. Second, for students who took the assessments online, we analyze only those respondents who demonstrated some effort by spending at least five minutes on the test.

Table 1 shows the proportions of students who are female, underrepresented minority (URM), first-generation collegegoers, and who are non-native English speakers in both the pre-pandemic (Spring or Fall 2019) and pandemic (Spring 2020) semesters. We cannot reject the hypotheses that these proportions are statistically equal between the pandemic and pre-pandemic semesters, except for finding a lower proportion of the first-generation students in the pandemic semester. It is possible that these students were more likely to withdraw from courses or college all together during the term. Any differences in these measures are addressed in our analyses through the inclusion of these demographic characteristics as controls in our models. We normalize the assessment scores by the mean and standard deviation of the pre-pandemic semester for each course. This allows us to pool the data from several courses and interpret effect sizes in terms of pre-pandemic standard deviations (SD).

Our survey of instructors asked about previous experience teaching online and whether they used particular teaching methods during the pandemic semester. Six of the seven classes were taught synchronously during the remote instruction period with lectures delivered to students in a Zoom meeting room. The seventh instructor pre-recorded lectures and spent the scheduled class time in Zoom answering student questions about the material.

In our analysis, we focus on two easily measured aspects of active learning pedagogy: use of polling software or "clickers" and explicit incorporation of peer interaction in the virtual classroom. Asking students to answer conceptual questions or solve problems during class has been shown to improve outcomes in in-person classes [e.g., *(4, 5)*] because it forces students to engage with the material and gives the instructor immediate feedback on what students have learned. We coded a course as using polling if the instructor polled students with at least two questions in all or all but one or two class meetings. Having students work together to answer challenging questions and engage in "peer instruction" has also been associated with positive student outcomes [e.g., *(6, 7)*]. We considered a course as using peer instruction if the instructor used at least two of the following strategies during the online portion of the pandemic semester: 1) classroom think-pair-share activities, 2) classroom small group activities, 3) encouraging students to work together outside class in pre-assigned small groups, and 4) allowing students to work together on exams. Our goal was to see whether online teaching experience or these two teaching techniques could potentially mitigate the negative effects of the pandemic in some courses.

We estimate three linear regression models for each of our two dependent variables: the standardized overall score on all assessment questions and the subscore based on the material that was taught remotely in the second portion of the Spring 2020 semester. Our first model estimates the effects of the pandemic separately for each of our seven study courses by including a course-specific fixed effect (μ_i) and separate course-specific effect for the pandemic semester (ϕ_{ip}):

$$y_{ips} = \mu_i + \phi_{ip} + \varepsilon_{ips}$$

The subscript *i* denotes the course, *p* is 1 during the pre-pandemic semester and 2 during the pandemic semester, and *s* indexes the student. The relative difference in average outcomes (pre-pandemic vs. pandemic) for each course is represented by the ϕ_{ip} term.

Our second model introduces a vector of controls for student demographic characteristics (Dem_{ips}) and interacts them with an indicator variable for the pandemic (d_p):

$$y_{ips} = \mu_i + \phi_{ip} + \beta_1 Dem_{ips} + \beta_2 Dem_{ips} \times d_p + \varepsilon_{ips}$$

β_1 represents the average effects of the demographic characteristics in the pre-pandemic semester while β_2 denotes the relative difference in these effects during pandemic semester.

We define our third model by replacing the course-specific pandemic effects with a single pandemic indicator variable (d_p) and interactions of that variable with a vector of three terms representing instructor and teaching characteristics (Ped_i):

$$y_{ips} = \mu_i + \alpha_1 d_p + \alpha_2 Ped_i \times d_p + \beta_1 Dem_{ips} + \beta_2 Dem_{ips} \times d_p + \varepsilon_{ips}$$

The three characteristics we include are whether the instructor has online teaching experience, whether the course included structured peer interaction in the classroom (e.g., working through problems in small groups), and whether the instructor used the common active learning technique of asking students to answer questions during class using polling software. In this model, α_1 is the average effect of the pandemic holding the instructor and teaching characteristics at zero, and α_2 is the average effect of each of these characteristics on learning during the pandemic semester relative to the non-pandemic semester.

We use Ordinary Least Squares (OLS) to obtain consistent point estimates of coefficients, but because the standard assumption of independence of error terms is violated in our context, we must use care in estimating our standard errors. Specifically, the unobservable shocks (ε_{ips}) are likely to be positively correlated for students in the same course. The conventional approach in this case is to calculate the *cluster-robust standard errors*, with each course serving as a cluster, but this method has been shown to perform poorly when the data contains a small (e.g., less than 30) number of clusters. Instead, we use the wild bootstrap method proposed in (*8*) to assess the statistical significance of estimated model coefficients because it allows us to conduct unbiased hypothesis tests even with a small number of clusters.

We standardize the assessment scores for each course using the pre-pandemic semester yielding the means of zero for the overall score and remote subscore shown in the first column of Table 1. In the pandemic semester, the overall score drops by 0.185 SD (p=0.015) while the remote subscore drops by 0.096 SD (p=0.181). This smaller and less precisely estimated effect is not altogether surprising, since these scores measure learning of topics taught closer to the administration of assessments, which potentially would be fresher in students' memory. Furthermore, at the institutions in this study, there was an extended break (up to three weeks) before the remote portion of the semester started. On the whole, these results suggest that student outcomes did suffer in the pandemic semester and the magnitudes of the declines in learning were not trivial.

The first two columns of Table 2 show that the effects of the pandemic on learning were very heterogeneous across courses. To illustrate, students in one course experienced a 0.836 SD decline in average overall scores, while students in another saw scores increase by 0.190 SD. All of these estimates differ significantly from zero (p-values shown in parentheses), and effects on the remote subscores are similarly varied.

In columns 3 and 4 of Table 2, we add controls for demographic characteristics in the models. This addition changes some of our course-specific estimates of the pandemic effect, but they remain very heterogenous and precisely estimated. The coefficients on the un-interacted demographic characteristics represent differences in learning in the pre-pandemic semester. They are mostly negative, replicating a common finding that female students and under-represented minorities (URM) often perform at lower levels than male or non-URM students in STEM courses [e.g., *(9,10)*]. We find that students who learned English as a second language (ESL) performed significantly worse than native English speakers on the material that was taught in the second portion of the course. Despite these direct effects, we see little evidence of interaction effects illustrating specific problems among these groups during the pandemic semester. Examining the interaction effects in the bottom rows of the table, we find very small and insignificant differences in performance in the pandemic semester for female and URM students relative to the pre-pandemic semester, and imprecise estimates of these differences for first generation and ESL status. Taken together, we see little evidence that students in different demographic groups were differentially affected by the pandemic.

Moving from course-specific to aggregate analysis, we estimate models in Table 3 that include a main effect for the pandemic semester, course-level fixed effects, demographic characteristics and variables representing each instructor's teaching experience and the teaching methods they used during the pandemic interacted with the pandemic indicator. Holding the demographic and instructor-level variables at zero, the pandemic and the emergency switch to remote instruction had a negative impact on student learning, especially for material that was taught during the remote portion of the semester where we see a statistically significant drop of 0.765 SD. That is, when instructors had no experience teaching online and did not include peer interaction or student polling when they taught remotely, our model predicts substantially lower scores in the pandemic semester relative to the pre-pandemic semester.

Consistent with results shown in Table 2, none of our demographic groups experienced significantly different effects of the pandemic relative to white or Asian male students that had at least one parent with a college degree and spoke English as their native language.

We find evidence that instructor experience and course pedagogy played important roles in ameliorating the potentially negative effects of the pandemic on learning. When the instructor had prior online teaching experience, student scores were significantly higher overall (0.611 SD, p=0.074) and for the remote material (0.625 SD, p=0.000). Students in classes with planned student peer interactions earned scores that were similar relative to students in other classes on the overall scores and 0.315 SD higher (p = 0.040) for the material taught remotely. We find no separate significant effect of polling students during class on student outcomes in the pandemic.

Our findings make us optimistic about future student learning outcomes even though we remain in a period of substantial online instruction. First, online teaching experience seems to matter, and during Spring 2020 most college faculty accumulated substantial experience. Second, we expected that disadvantaged groups would be further disadvantaged during the pandemic given their relative lack of support at home, but we found no statistical evidence of this concern. Third, we have shown that it is possible to incorporate peer interaction such as think-pair-share (*6*) or small group activities (*11*) into synchronous online courses, and that it was significantly associated with improved learning during the remotely taught portion of the semester.

References

1. University-Wide Active Learning Initiative. c2020. Cornell University: Office of the Provost; [accessed 2020 Oct 5]. https://provost.cornell.edu/leadership/vp-academic-innovation/active-learning-initiative/
2. W. K. Adams, C. E. Wieman, Development and validation of instruments to measure learning of expert-like thinking. *Int. J. Sci. Educ.*, 33(9), 1289-1312 (2011).
3. D. Xu, S S. Jaggars, Performance gaps between online and face-to-face courses: Differences across types of students and academic subject areas. *J. High. Educ.*, 85(5) (2014).
4. J. K. Knight, W. B. Wood, Teaching more by lecturing less. *Cell Biol. Educ.* 4(298) (2005).
5. R. A. Balaban, D. B. Gilleskie, U. Tran, A quantitative evaluation of the flipped classroom in a large lecture principles of economics course. *J. Econ. Educ.*, 47(4) (2016).
6. E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Saddle River, NJ, 1997)
7. C. H. Crouch, E. Mazur, Peer instruction: Ten years of experience and results. *Am. J. Phys.*, 69(970) (2001).
8. A. C. Cameron, J. B. Gelbach, D. L. Miller, Bootstrap-based improvements for inference with clustered errors, *Rev. Econ. Stat.*, 90(3), 414-427 (2008).
9. S. L. Eddy, S. E. Brownell, Beneath the numbers: A review of gender disparities in undergraduate education across science, technology, engineering, and math disciplines, *Phys. Rev. Phys. Educ. Res.*, 12, 020106 (2016).
10. T. G. Greene, C. N. Marti, K. McClenney, The effort—outcome Gap: Differences for African American and Hispanic community college students in student engagement and academic achievement, *J. High. Educ.*, 79(5) (2008).
11. S. A. Kalaian, R. M. Kasim, J. K. Nims, Effectiveness of small-group learning pedagogies in engineering and technology education: A meta-analysis, *J. Tech. Educ.*, 29(2) (2018).

Table 1. Descriptive Statistics: Student Learning Outcomes and Proportions of Demographic Groups.

	Pre-Pandemic Semesters		Pandemic Semester	
	Mean	Std. Dev.	Mean	Std. Dev.
Female	0.347	0.476	0.396	0.490
URM	0.130	0.337	0.111	0.315
First Generation	0.124	0.330	0.084+	0.278
ESL Speaker	0.269	0.444	0.240	0.428
Outcome (Overall)	0.000	1.000	-0.185*	1.112
Outcome (Remote)	0.000	1.000	-0.096	1.013
N of Observations	476		333	

Note: Significance tests of unconditional differences in means between pre-pandemic and pandemic semesters are shown using + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 2. Heterogeneous Effects of the Pandemic on Learning in Specific Courses

	(1) Overall		(2) Remote		(3) Overall		(4) Remote	
Course 1 × Pandemic	0.070**	(0.000)	0.017**	(0.000)	0.028	(0.574)	-0.123**	(0.002)
Course 2 × Pandemic	0.190**	(0.000)	0.310**	(0.000)	0.137	(0.208)	0.177*	(0.036)
Course 3 × Pandemic	-0.836**	(0.002)	-0.740**	(0.002)	-0.915**	(0.002)	-0.951**	(0.002)
Course 4 × Pandemic	-0.423**	(0.002)	-0.858**	(0.002)	-0.370**	(0.002)	-0.948**	(0.002)
Course 5 × Pandemic	-0.119**	(0.002)	-0.211**	(0.002)	-0.146	(0.252)	-0.360+	(0.074)
Course 6 × Pandemic	-0.360**	(0.002)	-0.149**	(0.002)	-0.446**	(0.002)	-0.335+	(0.074)
Course 7 × Pandemic	-0.625**	(0.002)	-0.353**	(0.002)	-0.678**	(0.002)	-0.497**	(0.002)
Female					-0.218+	(0.084)	-0.225	(0.120)
URM					-0.454**	(0.002)	-0.467**	(0.002)
FirstGen					-0.043	(0.892)	-0.096	(0.688)
ESL					0.016	(0.890)	-0.134*	(0.046)
Female × Pandemic					0.040	(0.666)	0.214	(0.160)
URM × Pandemic					-0.015	(0.962)	-0.0211	(0.936)
FirstGen × Pandemic					-0.315+	(0.078)	-0.0849	(0.830)
ESL × Pandemic					0.264	(0.378)	0.276	(0.122)
N of Observations	809		809		809		809	

Note: All equations include course-level fixed effects; *p*-values from wild bootstrap with course-level clustered standard errors hypothesis tests of zero effect in parentheses; + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 3. Effects of Pedagogy on Student Learning During the Pandemic.

	(1) Overall		(2) Remote	
	Coefficient	*p*-value	Coefficient	*p*-value
Pandemic	-0.641	(0.124)	-0.765**	(0.002)
Online Experience × Pandemic	0.611+	(0.074)	0.625**	(0.000)
Peer Interaction Online × Pandemic	0.047	(0.902)	0.315*	(0.040)
Student Polling × Pandemic	0.051	(0.936)	-0.025	(0.870)
Female	-0.210	(0.118)	-0.218	(0.136)
URM	-0.470**	(0.002)	-0.471**	(0.002)
First Gen	-0.043	(0.872)	-0.096	(0.706)
ESL	0.039	(0.652)	-0.123*	(0.046)
Female × Pandemic	0.030	(0.722)	0.204	(0.162)
URM × Pandemic	0.008	(0.940)	-0.030	(0.914)
First Gen × Pandemic	-0.247	(0.236)	-0.062	(0.846)
ESL × Pandemic	0.216	(0.510)	0.253	(0.136)
N of Observations	809		809	

Note: All equations include course-level fixed effects; *p*-values from wild bootstrap with course-level clustered standard errors hypothesis tests of zero effect in parentheses; + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$