



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



George Orlov^{a,*}, Douglas McKee^{a,*}, James Berry^b, Austin Boyle^c, Thomas DiCiccio^d, Tyler Ransom^e, Alex Rees-Jones^f, Jörg Stoye^a

^a Cornell University, Department of Economics, United States of America

^b University of Delaware, Department of Economics, United States of America

^c Pennsylvania State University, Department of Economics, United States of America

^d Cornell University, Department of Social Statistics, United States of America

^e University of Oklahoma, Department of Economics, United States of America

^f University of Pennsylvania, The Wharton School, Business Economics and Public Policy Department, United States of America

ARTICLE INFO

Article history:

Received 20 January 2021

Accepted 12 March 2021

Available online 16 March 2021

JEL classification:

A2

I2

Keywords:

COVID-19

Pedagogy

Student learning

ABSTRACT

We use unique data from seven intermediate economics courses taught at four R1 institutions to examine the effects of the COVID-19 pandemic on student learning. Because the same assessments of course knowledge mastery were administered across semesters, we can cleanly infer the impact of the unanticipated switch to remote teaching in Spring 2020. During the pandemic, total assessment scores declined by 0.2 standard deviations on average. However, we find substantial heterogeneity in learning outcomes across courses. Course instructors were surveyed about their pedagogy practices and our analysis suggests that prior online teaching experience and teaching methods that encouraged active engagement, such as the use of small group activities and projects, played an important role in mitigating this negative effect. In contrast, we find that student characteristics, including gender, race, and first-generation status, had no significant association with the decline in student performance in the pandemic semester.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

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. For many institutions, this transition was conducted on short notice, with little planning or prior experience to guide it. As of Spring 2021, many universities remain in this new instructional regime. 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 under these new conditions. It is also important to know whether specific groups of students were disproportionately harmed from the switch to remote learning, so that they can be provided with additional support.

Investigating how different aspects of teaching affect the learning of different types of students is 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 an in-person closed-book proctored exam to an open-book unproctored exam taken online). We circumvent this issue by analyzing data from seven intermediate-level economics courses in which student learning was measured using standard multiple-choice assessments with questions explicitly mapped to course learning goals. These assessments were developed at Cornell University as a part of the Active Learning Initiative,¹ following the procedure outlined in Adams and Wieman (2011) and administered as low-stakes tests at the end of each semester: The Intermediate Economics Skills Assessment – Microeconomics (IESA-Micro), the Economic Statistics Skills Assessment (ESSA), the Applied Econometrics Skills Assessment (AESA), and the Theory-based Econometrics Skills Assessment (TESA).

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

* Correspondence to: 109 Tower Rd., Uris Hall, Cornell University, Ithaca, NY, 14853, United States of America.

E-mail addresses: george.orlov@cornell.edu (G. Orlov), douglas.mckee@cornell.edu (D. McKee).

¹ See <https://provost.cornell.edu/leadership/vp-academic-innovation/active-learning-initiative/> for details.

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.² And third, we look at whether the use of specific teaching methods resulted in a more successful transition to remote teaching.

2. Data

Our data were collected during pre-pandemic (Spring or Fall 2019) and pandemic (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. After the Spring 2020 semester, instructors completed a survey regarding their teaching practices and material coverage before and during the pandemic. All but one course were taught by the same instructor in the pre-pandemic and pandemic semesters. Using the explicit mapping of assessment questions to course learning goals, we calculated a separate subscore for the material taught remotely in Spring 2020. We imposed two restrictions on our pooled analysis sample: First, students must have answered survey questions on gender, ethnicity, parental education, and non-native English speaker status. Second, for assessments administered 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 and pandemic semesters. We cannot reject the hypotheses that these proportions are equal between the pandemic and pre-pandemic semesters, except for finding a lower proportion of first-generation students in the pandemic semester. These students were potentially more likely to withdraw from courses or college altogether during the term. Any differences in these measures are addressed in our analyses through the inclusion of 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).

Instructors were surveyed about previous experience teaching online and teaching methods they employed during the pandemic semester. Six of the seven classes were taught synchronously during the remote instruction period with lectures delivered using Zoom. 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 during the online portion of the pandemic semester: use of polling software or “clickers” and 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 (Knight and Wood, 2005; Balaban et al., 2016) because it forces students to engage with the material and gives the instructor immediate feedback on what students have learned. Having students work together to answer challenging questions and engage in “peer instruction” has also been associated with positive student outcomes (Mazur, 1997; Crouch and Mazur, 2001). Here, we define

² 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 (Xu and Jaggars, 2014).

peer instruction as the use of at least two of the following strategies: 1) classroom think-pair-share activities,³ 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.

3. Estimation

We estimate three linear regression models for each of two dependent variables: the standardized overall score on all assessment questions and the subscore based on the material taught remotely during the pandemic semester. Our first model allows the effect of the pandemic to differ for each of our seven study courses by including a course-specific fixed effect (μ_i) and a separate course-specific effect for the pandemic semester (ϕ_{ip}):

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

where i denotes the course, p indicates the pre-pandemic or 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):

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

In our third model we replace the course-specific pandemic effects with a single pandemic indicator variable (d_p) and its interactions with a vector of three terms representing instructor and teaching characteristics (Ped_i):

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

Ped_i includes indicators for the instructor’s online teaching experience, the use of structured peer interaction in the classroom, and the use of the common active learning technique of asking students to answer questions during class using polling software.

We use OLS to obtain consistent point estimates of the coefficients. However, the unobservable shocks (ε_{ips}) are likely to be positively correlated for students within each course, causing OLS to yield biased standard errors. While the conventional approach in this case is to calculate the *cluster-robust standard errors*, with each course serving as a cluster, this method has been shown to perform poorly when the number of clusters is small (e.g., less than 30). Instead, we use the wild bootstrap method proposed in Cameron et al. (2008), which has been shown to have better performance with small numbers of clusters.

4. Results

Examining the assessment scores in Table 1, we see that in the pandemic semester, the overall score drops by 0.185SD ($p = 0.015$) while the remote subscore drops by 0.096SD ($p = 0.181$). A possible explanation for the discrepancy is that 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. Overall, these results suggest that student outcomes did suffer in the pandemic semester and the magnitudes of the declines in learning were not trivial.

³ In a think-pair-share exercise, students are first given a question and a few minutes to work on a solution individually. Next, they work with a partner to discuss their approach and try to come to a consensus answer (Lyman, 1987; King, 1993).

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)		(2)		(3)		(4)	
	Overall		Remote		Overall		Remote	
	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
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 in parentheses are computed from hypothesis tests of zero effect using wild bootstrap with course-level clustered standard errors; + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

The first two columns of Table 2 show that the effects of the pandemic on learning were very heterogeneous across courses, with effects ranging from a 0.836SD decline in average overall scores to a 0.190SD increase. All of these estimates differ significantly from zero, 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 heterogeneous 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 (Eddy and Brownell, 2016; Greene et al., 2008). 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. 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, Table 3 shows estimates for models that introduce instructor's teaching experience and the teaching methods 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 0.765SD. 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 (0.611SD, $p = 0.074$) and for the remote material (0.625SD, $p = 0.000$). Students in classes with planned student peer interactions earned scores that were similar relative to students in other classes on the overall scores and 0.315SD 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.

Table 3
Effects of pedagogy on student learning during the pandemic.

	(1)		(2)	
	Overall		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 in parentheses are computed from hypothesis tests of zero effect using wild bootstrap with course-level clustered standard errors; + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

5. Conclusion

Our findings make us optimistic about future student learning outcomes even though we remain in a period of substantial online instruction for three reasons. First, online teaching experience seems to matter, and during 2020 many college faculty accumulated some experience. Second, we expected that disadvantaged groups would be further disadvantaged during the pandemic, 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 (Mazur, 1997) or small group activities (Kalaian et al., 2018) into synchronous online courses, and that it was significantly associated with improved learning during the remotely taught portion of the semester.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

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.

References

- Adams, W.K., Wieman, C.E., 2011. Development and validation of instruments to measure learning of expert-like thinking. *Int. J. Sci. Educ.* 33 (9), 1289–1312.
- Balaban, R.A., Gilleskie, D.B., Tran, U., 2016. A quantitative evaluation of the flipped classroom in a large lecture principles of economics course. *J. Econ. Educ.* 47 (4).
- Cameron, A.C., Gelbach, J.B., Miller, D.L., 2008. Bootstrap-based improvements for inference with clustered errors. *Rev. Econ. Stat.* 90 (3), 414–427.
- Crouch, C.H., Mazur, E., 2001. Peer instruction: Ten years of experience and results. *Amer. J. Phys.* 69 (970).
- Eddy, S.L., Brownell, S.E., 2016. 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.
- Greene, T.G., Marti, C.N., McClenney, K., 2008. The effort-outcome Gap: Differences for African American and Hispanic community college students in student engagement and academic achievement. *J. High. Educ.* 79 (5).
- Kalaian, S.A., Kasim, R.M., Nims, J.K., 2018. Effectiveness of small-group learning pedagogies in engineering and technology education: A meta-analysis. *J. Tech. Educ.* 29 (2).
- King, A., 1993. From sage on the stage to guide on the side. *Coll. Teach.* 41 (1).
- Knight, J.K., Wood, W.B., 2005. Teaching more by lecturing less. *Cell Biol. Educ.* 4 (298).
- Lyman, F., 1987. Think-Pair-Share: An expanding teaching technique. *MAA-CIE Coop. News* 1.
- Mazur, E., 1997. *Peer Instruction: A User's Manual*. Prentice Hall, Saddle River, NJ.
- Xu, D., Jaggars, S.S., 2014. Performance gaps between online and face-to-face courses: Differences across types of students and academic subject areas. *J. High. Educ.* 85 (5).