Appendix for "Zooming to Class?: Experimental Evidence on College Students' Online Learning during COVID-19"

Michael S. Kofoed[*], Lucas Gebhart[†], Dallas Gilmore[‡], and Ryan Moschitto[§]

July 12, 2024

[*]Michael Kofoed is an assistant professor at the University of Tennessee, Knoxville and Research Fellow at the IZA: 916 Volunteer Boulevard, Knoxville, Tennessee 37916, email: *mkofoed1@utk.edu.*

[†]Lucas Gebhart is an officer in the United States Army

[‡]Dallas Gilmore is an officer in the United States Army.

[§]Ryan Moschitto is an office in the United States Army

1 Comparison of West Point Cadets to Civilian Colleges

To address concerns about external validity, we compare West Point to other civilian institutions. Table A1 shows descriptive statistics for the 75th and 25th SAT Math scores and the percent of the student body that is female, black, Hispanic, or Asian. We contrast our sample with the data from the Integrated Post-secondary Education Data System (IPEDS) to better understand the type of student that is similar to those at West Point Integrated Postsecondary Education Database (2022). First, we compare our sample to data from public, flagship institutions.[1] In Column (2), we consider a school a "flagship" institution if the university boasts the highest 75th percentile SAT Math school among public universities in a given state. Here, we find that West Point students have slightly higher SAT scores, lower number of female students, a higher percentage of black students, a lower percentage of Hispanic students and higher percentages of Asian students.

In Column (3), we show descriptive statistics from the "Top 15" national universities[2] as determined by the U.S. News and World Report rankings. We find that West Point students have lower SAT scores than top national universities. West Point cadets are also less likely to be female, more likely to be black, and less likely to be Hispanic or Asian. These results indicate that West Point students are slightly better prepared for college than public, flagship university students but less prepared than those at "Top 15" national universities.

2 Balance Table

We show covariate balance and test the efficacy of the experiment by regressing whether a student was in an online class on a number of covariates, instructor fixed effects, class day fixed effects, and time of day fixed effects. Table A2 shows the results from these regression estimates. First, we separately regress assignment to an online section on each of our covariates. Columns (1) though (7) show results for being female, black, Hispanic, Asian, or an NCAA athlete; having attended the prep school; and being prior enlisted. All of these coefficients are small and statistically insignificant, evidence that students were randomly assigned to an online section. We also show F-statistics and their corresponding *p*-values to test the efficacy and find no indication of statistically significant imbalance.

In Column (8), we jointly regress being in an online class section on all of our covariates. We find that NCAA athletes were less likely to be in an online class section, a result that is marginally statistically significant. Thus, we add time of day fixed effects to show that within a class hour, West Point did randomly assign students to either online or in person. However, the F-stat *p*-value for this specification is only 0.512. Column (9) shows results

[1] The median SAT Math score is 670 which is similar to engineering focused universities such as Texas A&M, Purdue University, or Virginia Tech

[2] These institutions include keep if Princeton University, Columbia University, Harvard University, Massachusetts Institute of Technology, Yale University, Stanford University, University of Chicago, University of Pennsylvania, California Institute of Technology, Duke University, Johns Hopkins University, Northwestern University, Dartmouth College, Brown University, Vanderbilt University

for the same regression but adding instructor fixed effects and finds that prep school students were more likely to be in an online section, while prior enlisted students were less likely to be in an online section. However, these results are only marginally significant with an F-stat p-value of 0.552. Column (10) adds class day fixed effects to the model, with NCAA athlete and prior enlisted being marginally statistically significant with an F-stat p-value of 0.499. Finally, Column (11) includes all covariates and adds time of day fixed effects. Conditional on time of day, all of our covariates balance with an F-stat p-value of 0.578.

3 Quartile Regression

Next, we estimate a quantile regression to show differences in the effect by final grade. We estimate specifications for the 10th, 25th, 50th, 75th, and 90th quantiles. Table A4 displays these results. Column (1) is the overall effect for comparison. Column (2) finds that online students in the bottom 10th percentile received grades that were 0.090 standard deviations less than similar in-person students. However, Columns (3) and (4) show that for students at the 25th and 50th percentile, online coursework reduced their final grades by 0.225 and 0.351 standard deviations, respectively, results that are statistically significant. As we move up the final grade distribution, the achievement gap between online and in-person coursework begins to close. Columns (5) and (6) show achievement gaps of 0.115 and 0.194 standard deviations, results that are either statistically insignificant or marginally significant depending on the specification.

4 Mechanisms

In this section, explore potential mechanisms that could help to explain our result. First, we examine if faculty teaching experience is driving our result. Since most of our faculty are rotating military faculty, we compare those who have never taught before to those who have been at West Point for at least one year.

4.1 Mechanisms: Faculty Experience

We examine whether faculty teaching experience played a role in determining online class outcomes. The West Point faculty model is unique in that approximately half the of instructors are rotating, active duty military faculty who temporarily leave the operational Army, attend graduate school (usually earning a terminal master's degree such as an MBA/MPA, though a PhD is possible), and teach at West Point for three years before returning to the operational Army.[3] One possible concern is that instructor experience could be correlated with better outcomes in an online environment and teaching experience could explain our results as opposed to the online environment. To address this

[3]Only one tenured/tenure track faculty member with a PhD participated in our experience. All other instructors were rotating military faculty.

concern, we estimate our regression model with students taught by first-year and experienced faculty separately. To see if these treatment effects are statistically different from each other, we estimate a model that interacts the indicator variable for being in an online class with whether a first-year faculty member was the instructor. While online instruction was new at West Point, experienced faculty members did have switch online halfway through the Spring 2020 semester. These faculty also had the entire summer to prepare to teach online in the fall. New rotating military faculty had no experience teaching online.

Table A5 displays the results for instructor experience. Column (1) shows the results for new instructors only. We find that students taught in online courses by first-year instructors received a grade that was 0.256 standard deviations lower than students in in-person classes, a result that is statistically significant. Column (2) shows a similar albeit less precise estimate of 0.249 standard deviations for being in an online class. Column (3) shows the interaction between being assigned both an online class and a new faculty member and shows that the difference between new and experienced faculty members is not statistically significant. Thus we find that the grade penalty for being in the online environment is similar across faculty experience.

4.2 Mechanisms: Class Size Differences

One concern with our experiment design is that West Point limited in-person class sizes such that the average in-person class had twelve students while the average online class had seventeen. Thus, it is difficult to untangle the online result from any gains to being in a smaller class since both characteristics are wrapped up in the same treatment arm. However, since the average difference in class size is five students, previous literature on returns to class size in higher education can help us bound what part of our estimate can be attributed to class sizes as opposed to being online.

Compared to class size effects in the K-12 literature,[4] effects in higher education seem to be more muted. Bandiera, Larcinese and Rasul (2010) use a panel of college classes in the UK and find that a one standard deviation increase (33.2 students) decreases final grades by -0.108 standard deviations. Thus the average class size effect in our setting would be around 0.016 standard deviations or account for about eight percent of our estimates.

In a similar setting, Kara, Tonin and Vlassopoulos (2021) estimate that a one standard deviation increase (around fifteen students) reduced course grades by 0.08 standard deviations. Thus, in our experiment, a six student increase in class size reduces student grades by 0.032 standard deviations or around sixteen percent of our result. De Giorgi, Pellizzari and Woolston (2012) find similar results where a one standard deviation increase (around eighteen students) in class sizes reduces course grades by 0.10 standard deviations.

[4]The Tennessee STAR experiment is the most famous example of randomly sorting students to different class sizes in the K-12 literature. Krueger (1999) and Krueger and Whitmore (2001) find positive results in both short and long-terms for lower class sizes. However, they could potentially only explain a portion of our result and apply to Kindergartners not college students.

Finally, regarding class size effects in online classes, Bettinger et al. (2017) leverage a quasi-random field experiment at a large, online, for-profit university to measure the effects of class sizes on academic achievement. The university sorted students into classes of either 31 or 34 students. The authors find no economically significant effect of increasing class size. While the difference in the number of students is not large, it is similar to many of the class sections in our setting.

We test the concern that class sizes may be driving our results by estimating a number of alternative specifications. Since West Point randomly assigns students to class sections (even within day and hour), the class sizes are also exogenously determined. We would be concerned about our results if it appeared that class size explained more (or all) of the variation than class modality. Table A6 shows specifications similar to the main result, but where we add the class size of the section. Column (2) shows that the addition of class size to the regression increases the penalty for being in an online class to 0.505 standard deviations. We also find that an additional class member increases a student's final grad by 0.055 standard deviations. However, there could be non-linearities in the class size effect given the already small classes[5]. To account for these non-linearities, we add the class size and its square. Column (3) shows that the online penalty decreases slightly to 0.438 standard deviations. The sign on the class size flips but remains small with a small positive relationship for the square. However, neither of these class size coefficients are statistically significant.

Finally, we interact the indicator for being in an online class with the number of students in that class section. Column (3) shows the results from this specification. We find that class size in an online class seems to increase final grades of students compared to in-person classes, but this result is not precise.

While the online sections did not have the same capacity constraints as the in-person classes, given the random assignment, some online class sizes did overlap with some in-person sections. While the average in-person section was twelve students, some sections had up to fourteen while some online sections had as few as thirteen. While not perfect, comparing classes of similar size should eliminate any difference in outcomes if class size is the main driver of our results.

Table A6 shows results from this exercise. First, we compare class sections between twelve and fifteen students; essentially cutting the class size difference in half. In Column (5), We find that the grade penalty for being in an online class actually increases to 0.418 standard deviations; a result that is also statistically significant. Next, we tighten the comparison group by only considering class sections with enrollments between twelve and fourteen students. In Column (6), we find an online penalty of 0.570 standard deviations. However, this result is no longer statistically significant because of the considerable reduction in sample size. Finally, in Column (7), we find a similar result when comparing classes with enrollments between twelve and fourteen students. While this evidence does not rule out a role for class size effects, previous literature and considering class sections of similar sizes show that class

[5]For example, an additional student in a class of fifteen may have more of an effect than an additional student in a lecture hall of 100 students.

size may have only a limited role in explaining our results.

4.3 Follow-On Results

Finally, we examine some, albeit limited, follow-on effects from our experiment. We consider whether a student switched their major to economics and academic performance in upper-division economics classes that students took in Spring Semester 2021. Table A7 show the results from each of the models. The results should be taken with some caution since West Point students declare their majors during their majors in their first year and face some switching costs for changing. In our sample, only nine students changed their majors to economics (out of 551). Also, only majors take upper-division economics courses and there were only twenty-four in our sample. Thus these estimates are under-powered, however, we discuss them with these caveats.

First, we estimate our model with whether a student switches to economics as a major. We find a small negative effect of 0.4 percentage points for online class sections. This result is not statistically significant, but since 1.63 percent of our sample switches to econ, this result is somewhat sizeable relatively. Next, we look at the final grades for economics majors in our experiment in the next follow on course "Mathematics for Economics". Since only twenty-one students that participated in our experiment when on to take this course, these results should be considered with caveat and are not statistically significant. We find that students in an online course earned a final grade that was 1.722 standard deviations less than their colleagues in a face to face class. Some of our students took multiple follow-on courses during the next semester.[6]. We use the average, final grade earned as a dependent variable. We find a reduction of 0.394 standard deviations. These results are very under-powered, but do indicate a possible reduction in longer run outcomes from being in an online course.

[6]These courses include Econometrics, Financial Statement and Firm Analysis, Game Theory, or Intermediate Microeconomics

References

Bandiera, Oriana, Valentino Larcinese, and Imran Rasul. 2010. "Heterogeneous Class Size Effects: New Evidence from a Panel of University Students." *The Economic Journal*, 120(549): 1365–1398. 3

Bettinger, Eric P., Susanna Loeb Christopher Doss, Aaron Rogers, and Eric S. Taylor. 2017. "The Effects of Class Size in Online College Courses: Experimental Evidence." *Economics of Education Review*, 58: 68–85. 4

De Giorgi, Giacomo, Michele Pellizzari, and William Gui Woolston. 2012. "Class Size And Class Heterogeneity." *Journal of the European Economic Association, European Economic Association*, 10(4): 795–830. 3

Integrated Postsecondary Education Database. 2022. ""Fall Enrollment" and "Admissions and Test Scores"." *https://nces.ed.gov/ipeds/datacenter/cds.aspx?sid=17f0e3da-7b89-4ad5-b78e-230c60a938e9&rtid=5*. 1

Kara, Elif, Micro Tonin, and Michael Vlassopoulos. 2021. "Class Size Effects in Higher Education: Differences Across STEM and Non-STEM Fields." *Economics of Education Review*, 82: 1–13. 3

Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions." *The Quarterly Journal of Economics*, 114(2): 497–532. 3

Krueger, Alan B., and Diane M. Whitmore. 2001. "The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR." *Economic Journal*, 111(468): 1–28. 3

Table A1: Summary Statistics Comparing Full Sample to Public Flagship and Top 15 Universities

	(1) Full Sample mean/sd	(2) Public Flagship mean/sd	(3) Difference diff/se	(4) Top 15 Univ. mean/sd	(5) Difference diff/se
SAT Math 75th Percentile	730	707.03	22.97**	795.02	-65.02***
	(70.00)	(50.06)	[9.942]	(6.256)	[18.09]
SAT Math 25th Percentile	620	594.70	25.30**	726.46	-106.46***
	(70.00)	(55.86)	[9.997]	(16.17)	[18.10]
Female	0.230	0.504	-0.274***	0.494	0.264**
	(0.422)	(0.054)	[0.059]	(0.043)	[0.109]
Black	0.140	0.051	0.089*	0.056	0.084
	(0.347)	(0.041)	[0.048]	(0.015)	[0.090]
Hispanic	0.033	0.086	-0.053**	0.083	0.050
	(0.178)	(0.068)	[0.025]	(0.090)	[0.046]
Asian	0.056	0.095	-0.039	0.141	0.085
	(0.231)	(0.073)	[0.032]	(0.019)	[0.060]
Observations	551	52		15	

This tables shows differences student demographics between West Point, state public flagship universities, and the U.S. News and World Report Top 15 Universities. State public flagship universities are determined by the public institution with the highest SAT math score in a given state. Means are weighed by student enrollment and displayed in the first row with standard deviations in parentheses. Columns (3) and (5) show the difference between West Point and a given category with standard errors below in brackets. Statistical significance levels are as follows: * for $p<0.10$, ** for $p<0.05$, and *** for $p<0.01$.

Table A2: Balance Table for Online Instruction

	(1) Online	(2) Online	(3) Online	(4) Online	(5) Online	(6) Online	(7) Online	(8) Online	(9) Online	(10) Online	(11) Online	(12) Online
Female	-0.017								0.003	0.007	0.009	0.002
	(0.050)								(0.051)	(0.051)	(0.051)	(0.050)
Black		-0.001							-0.027	-0.028	-0.033	-0.027
		(0.060)							(0.065)	(0.065)	(0.066)	(0.064)
Hispanic			-0.115						-0.130	-0.120	-0.115	-0.166
			(0.120)						(0.119)	(0.124)	(0.122)	(0.124)
Asian				-0.067					-0.083	-0.071	-0.072	-0.028
				(0.092)					(0.094)	(0.099)	(0.098)	(0.096)
NCAA Athlete					-0.069				-0.093*	-0.087*	-0.098*	-0.076
					(0.046)				(0.050)	(0.051)	(0.050)	(0.049)
Prep School						0.004			0.138	0.156	0.147	0.151
						(0.060)			(0.110)	(0.108)	(0.105)	(0.110)
Prior Enlisted							-0.037		-0.157	-0.170*	-0.164*	-0.174*
							(0.056)		(0.101)	(0.097)	(0.093)	(0.099)
CEER Score								-0.000	-0.000	-0.000	-0.000	-0.000
								(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Instructor FE	No	No	No	No	No	No	No	No	No	Yes	Yes	Yes
Class Day FEs	No	No	No	No	No	No	No	No	No	No	Yes	Yes
Time of Day FEs	No	No	No	No	No	No	No	No	No	No	No	Yes
F_Stat	0.119	0.001	0.926	0.530	2.238	0.005	0.441	0.081	0.914	0.876	0.964	0.952
F_p_value	0.730	0.981	0.336	0.467	0.135	0.941	0.507	0.776	0.505	0.536	0.463	0.473

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table is an additional measure of covariate balance. We show that, jointly, neither instructor fixed effects, class day fixed effects, time of day fixed effects, exogenous controls, or combinations of these variables can explain whether West Point assigns a student to an online class. Possible exceptions include whether a student is an NCAA athlete or prior enlisted; however, these are only marginally significant.

Table A3: Main Effects for Online Instruction: Clustered Standard Errors

	(1)	(2)	(3)	(4)	(5)
	Final Grade	Final Grade	Final Grade	Final Grade	Final Grade
Online	-0.236***	-0.220***	-0.223***	-0.218***	-0.215***
	(0.072)	(0.063)	(0.063)	(0.055)	(0.050)
Instructor FE	No	Yes	Yes	Yes	Yes
Class Day FEs	No	No	Yes	Yes	Yes
Time of Day FEs	No	No	No	Yes	Yes
Exog. Controls	No	No	No	No	Yes
Observations	551	551	551	551	551
R^2	0.013	0.026	0.026	0.034	0.173

Class section level clustered standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table shows the main result for the effect of online learning on college student academic achievement. Instructor FE indicate instructor fixed effects. Exogenous controls include whether a student is female, black, Hispanic, Asian, or an NCAA athlete; attended the US Military Academy Preparatory School; or was prior enlisted. We also add instructor fixed effects, class day fixed effects, and time of day fixed effects to control for the level of random assignment.

Table A4: Quantile Regression for Final Course Grade

	(1)	(2)	(3)	(4)	(5)	(6)
	Final Grade					
	Overall	Q.10	Q.25	Q.50	Q.75	Q.90
Online	-0.215**	-0.090	-0.226**	-0.351***	-0.115	-0.194*
	(0.084)	(0.173)	(0.111)	(0.119)	(0.091)	(0.112)
Instructor FE	Yes	Yes	Yes	Yes	Yes	Yes
Class Day FEs	Yes	Yes	Yes	Yes	Yes	Yes
Time of Day FEs	Yes	Yes	Yes	Yes	Yes	Yes
Exog. Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	551	551	551	551	551	551
R^2	0.173	0.126	0.127	0.130	0.103	0.098
Robust SEs p-values	0.011	0.603	0.042	0.003	0.209	0.084
Wild Bootstrapped SEs p-values	0.013	0.009	0.016	0.018	0.014	0.009

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table shows the result for the effect of online learning on college student academic achievement across the distribution of final exam grades. We estimate a quartile regression for the 10th, 25th, 50th, 75th, and 90th percentiles. Exogenous controls include whether a student is female, black, Hispanic, Asian, an NCAA athlete; attended the US Military Academy Preparatory School; or was prior enlisted. We also add instructor fixed effects, class day fixed effects, and time of day fixed effects to control for the level of random assignment.

Table A5: Effects of Online Instruction by Faculty Experience

	(1)	(2)	(3)	(4)
			Final Grade	
	Overall	New Instructor	Experienced Instructor	Interact-New Instructor
Online	-0.215**	-0.256**	-0.249	-0.129
	(0.084)	(0.118)	(0.181)	(0.137)
New Instructor				0.132
				(0.264)
Online × New Instructor				-0.143
				(0.179)
Instructor FE	Yes	Yes	Yes	Yes
Class Day FEs	Yes	Yes	Yes	Yes
Time of Day FEs	Yes	Yes	Yes	Yes
Exog. Controls	Yes	Yes	Yes	Yes
Observations	551	310	241	551
R^2	0.173	0.170	0.204	0.174
Robust SEs p-values	0.011	0.031	0.169	0.348
Wild Bootstrapped SEs p-values	0.013	0.024	0.185	0.371

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table shows the result for the effect of online learning on college student academic achievement by instructor experience. New Instructor is an instructor in their first year, while Experienced Instructor is one who is beyond their first year. Instructor FE indicate instructor fixed effects. Exogenous controls include whether a student is female, black, Hispanic, Asian, or an NCAA athlete; attended the US Military Academy Preparatory School; or was prior enlisted. We also add instructor fixed effects, class day fixed effects, and time of day fixed effects to control for the level of random assignment.

Table A6: Role of Class Size

	(1) Overall	(2) Full Sample	(3) Full Sample	(4) Full Sample	(5) 12-15 Students	(6) 12-14 Students	(7) 12-13 Students
Online	-0.215**	-0.505***	-0.438*	-2.071	-0.418***	-0.570	-0.575
	(0.084)	(0.160)	(0.256)	(1.621)	(0.140)	(0.470)	(0.473)
Class Size		0.055**	-0.050	-0.066			
		(0.027)	(0.301)	(0.132)			
$ClassSize^2$			0.003				
			(0.009)				
Online × Class Size				0.127			
				(0.132)			
Instructor FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class Day FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time of Day FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exog. Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
obs	551	551	551	551	291	217	203
r2	0.173	0.179	0.180	0.181	0.199	0.217	0.207
Robust p-value	0.011	0.002	0.088	0.202	0.003	0.227	0.227

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table shows the main result for the effect of online learning on college student academic achievement. Instructor FE indicate instructor fixed effects. Exogenous controls include whether a student is female, black, Hispanic, Asian, or an NCAA athlete; attended the US Military Academy Preparatory School; or was prior enlisted. We also add instructor fixed effects, class day fixed effects, and time of day fixed effects to control for the level of random assignment.

Table A7: Follow-On Effects of Online Learning

	(1) Final Grade	(2) Switch to Econ	(3) Math for Econ Grades	(4) Mean Follow on Course Grade
Online	-0.215**	-0.004	-1.722	-0.394
	(0.084)	(0.005)	(1.515)	(0.871)
Instructor FE	Yes	Yes	Yes	Yes
Class Day FEs	Yes	Yes	Yes	Yes
Time of Day FEs	Yes	Yes	Yes	Yes
Exog. Controls	Yes	Yes	Yes	Yes
Observations	551	525	21	24
R^2	0.173	0.062	0.986	0.985
Robust SEs p-values	0.011	0.405	0.373	0.682
Wild Bootstrapped SEs p-values	0.009	0.661	0.280	0.762

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table shows the main result for the effect of online learning on college student academic achievement. Instructor FE indicate instructor fixed effects. Exogenous controls include whether a student is female, black, Hispanic, Asian, or an NCAA athlete; attended the US Military Academy Preparatory School; or was prior enlisted. We also add instructor fixed effects, class day fixed effects, and time of day fixed effects to control for the level of random assignment.