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

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

July 7, 2021

Abstract

COVID-19 shifted schools and colleges to online instruction with little causal evidence of outcomes. In the fall of 2020, we randomized 551 West Point students in a required introductory economics course across twelve instructors to either an online or in-person class. Final grades for online students dropped by 0.215 standard deviations, a result apparent in both assignments and exams and largest for academically at-risk students. A post-course survey finds that online students struggled to concentrate in class and felt less connected to their instructors and peers. We find that the shift to online education had negative results for learning.

*The authors would like to thank Russell Lachance for data and assistance with the randomization. We thank Cindy Jebb, Rachel Sondheimer, Suzanne Nielsen, Heidi Demarest, and Spencer Clouatre for helping us obtain permission for this experiment. We appreciate the helpful comments of Susan Carter, Sarah Cohodes, Eric Eide, Christopher Fawson, David Lyle, Lars Lefgen, Jonathan Meer, Brad Mortensen, Richard Patterson, Joshua Price, Mark Showalter, Carl Wojtaszek, Michael Yankovich, and Ulf Zölitz. We are also grateful for participants at conference presentations including the Association for Education Finance and Policy. We are very grateful to the many Principles of Economics instructors who agreed to teach half of their course load online and half in person to ensure the success of our experiment. The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, and the Department of Defense. This experiment was pre-registered in the American Economic Association RCT Registry as AEARCTR-0006969.

†Michael Kofoed is an associate professor at the United States Military Academy and research fellow at the IZA: 607 Cullum Road, West Point, New York, USA, 10996; tel: (845) 938-4110, email: *michael.kofoed@westpoint.edu.*

‡Lucas Gebhart is an instructor of economics at the United States Military Academy; tel: (845) 938-3376, email: *Lucas.Gebhart@westpoint.edu.*

§Dallas Gilmore is an instructor of economics at the United States Military Academy; tel: (845) 938-2857, email: *Dallas.Gilmore@westpoint.edu*.

¶Ryan Moschitto is an instructor of economics at the United States Military Academy; tel: (845) 938-0017, email: *Ryan.Moschitto@westpoint.edu.*

1 Introduction

The COVID-19 pandemic posed difficult challenges to many colleges and universities. Of primary concern was how to deliver instruction while maintaining social distancing and other safety measures essential to slowing the spread of infection. In March of 2020, many colleges pivoted to online instruction to finish the semester[1] and continued with online learning during the next academic year. However, the following fall semester was a challenge for many college administrators who had to choose between in-person, online, or hybrid alternatives in the face of uncertainty from the virus. Colleges that reopened either offered students the decision to attend in-person classes, watch live lectures online, or select a hybrid of the two. As the fall 2020 semester progressed, spiking infections caused some colleges to teach completely online. Thus, self-selection of students, instructors, and (in many cases) entire institutions into online versus in-person classrooms makes studying the causal effects of online learning during the pandemic difficult.

In this paper, we conduct a randomized controlled trial at the United States Military Academy at West Point in a required Principles of Economics course to estimate differences in student outcomes between online and in-person instruction. In a non-COVID-19 environment, West Point students (along with their peers at the Air Force Academy and Naval Academy) have little control over their academic schedules and peer networks. West Point also randomly assigns students to academic instructors (Carrell, Page and West, 2010; Carrell and West, 2010; Mansour et al., Forthcoming), the semester they take a general education course (Patterson, Pope and Feudo, 2019), class hours (Carrell, Maghakian and West, 2011; Haggag et al., 2019), roommates (Jones and Kofoed, 2020), social networks called companies (Lyle, 2007, 2009; Brady, Insler and Rahman, 2017), final exam periods, and military mentors (Kofoed and mcGovney, 2019).

We randomized 550 students across 12 instructors in 38 class sections. In our experiment, the syllabus, graded events, homework assignments, and exams[2] were identical for online and in-person classes. We also offered both online and in-person class sections for each class hour the course was available, and we randomly assigned students within the hour into each modality.[3] In this study, we find that online education lowered a student's final grade by 0.215 standard deviations, a statistically significant result driven by the students with below median academic ability.

Online education has been controversial since its inception. Over the last few decades, many technologists have predicted that online education would disrupt traditional markets and education delivery (Christensen and Eyring, 2011; Crow and Dabars, 2015). Proponents argue that universities can use online classes to expand their student base beyond geographic locations, lower costs of instruction, and increase access for non-traditional students.

[1] Mangrum and Niekamp (2020) leverage the timing of spring break travel to show that the suspension of in-person instruction was effective for slowing the spread of COVID-19 infection among college students and the local areas surrounding college campuses.

[2] There were two versions of the exam, one for each day the exam was offered. We included class day fixed effects to account for this difference.

[3] Offering the exam online for both modalities would also ensure that if students participated in academic dishonesty (as documented by Bilen and Matros, 2021), then there would not be differential opportunities for this behavior between treatment and control.. In our setting there was equal opportunity for this behavior across treatment and control.

For example, Deming et al. (2015) show that colleges and universities decrease tuition rates as they increase their online course offerings, with the largest tuition declines seen at public and for-profit institutions. Cowen and Tabarrok (2014) use an industrial organization framework to argue that online education can reduce the costs of education by spreading out the fixed cost of labor-intensive instruction over a greater student body. Deming, Lovenheim and Patterson (2020) show that online higher education does compete against traditional brick-and-mortar institutions by reducing enrollments at traditional colleges and inducing these colleges to increase per-student spending.

While online instruction seems to offer colleges a cost-saving mechanism, the educational benefits are not as clear. The main difficulty of calculating the effect of online courses is selection bias from students self-sorting into modality of instruction; students who enroll in online courses tend to be older, more likely to work during the semester, and live farther away from campus (Coates et al., 2004; Dutton, Dutton and Perry, 2002). Since different types of students sort into online courses, any estimated gap between the two modalities would be biased, rendering it unclear if online courses are as effective as in-person instruction. Coates et al. (2004) use a two-step estimator to show that failure to address this bias causes outcomes to look similar and may explain null effects in previous descriptive work (Russell, 1999).

There have been a few randomized control trials that measure the effect of online instruction including Figlio, Rush and Yin (2013), Joyce et al. (2015), and Alpert, Couch and Harmon (2016). These experiments were limited to one instructor in one course where researchers recruited students to participate using grade incentives. The results in each study were mixed, finding that online education either reduced student grades early in the semester (Joyce et al., 2015), in-person only courses boosted grades for some students (i.e., male, Hispanic, or low-achieving students) (Figlio, Rush and Yin, 2013), or strictly online classes hurt students compared to peers in a hybrid format (Alpert, Couch and Harmon, 2016). Cacault et al. (Forthcoming) conduct an experiment at a Swiss university that covers multiple instructors in multiple courses where they randomly assign access to video lectures for some weeks during the semester and remove access for other weeks. They find that access to streamed lectures increases grades for high-ability students but decreases outcomes for less able students. They also find that streaming decreases the incentive to attend in-person lectures.

Using an instrumental variables approach as opposed to a randomized control trial, Bettinger et al. (2017) use data from more than two million students at a large for-profit university that offered both online and in-person classes. The authors calculate the distance from a student's residence to the nearest brick-and-mortar campus as an instrument to address endogeneity of student sorting. The authors find that students who took similar online classes earned lower course grades and were less likely to graduate.

Finally, there is an expanding literature regarding technology in the classroom (even if technology does not subsume the entire class experience). Carter, Greenberg and Walker (2017) conduct a randomized control trial at West Point where instructors either required students to use a laptop for note taking, permitted tablets that students laid faceup on their desks, or banned laptops completely. The authors find a negative effect in both treatment arms. Setren et al. (Forthcoming) use a similar experiment regarding a flipped classroom environment to show that mathematics students in the flipped classroom[4] had modest, short-term gains. However, these gains were among white, male, and high-ability students, thus increasing inequality. Patterson and Patterson (2017) use a natural experiment where some students were required to bring in their laptops on certain days at a liberal arts college. The authors find that students saw their grades decline in classes taught on the same day as a class that required laptops.

Our paper contributes to the literature in several ways. First, our randomized control trial does not depend on volunteers since Principles of Economics is a required course, eliminating the concern about selection bias into the experiment. Also, West Point randomizes the semester[5] in which a student takes any of their required classes, meaning there is no self-selection into or out of this economics course in a particular semester. Next, we randomized students across twelve instructors with each instructor teaching a balanced slate of online and in-person class sections. This feature allows us to control for instructor attributes like experience and comfort with either teaching modality. Finally, since this paper uses a randomized control trial setting, we do not rely on more strict identification assumptions like those in an instrumental variables design.

The rest of this paper proceeds as follows. Section II describes the institutional setting, Section III describes the random assignment and the experimental setting, and Section IV describes the econometric models. Section V discusses the results, Section VI considers policy implications and limitations, and Section VII concludes.

2 Institutional Details

The US Military Academy at West Point, New York, is a highly selective, public liberal arts college and the nation's oldest military academy. West Point's primary mission is to educate future officers who go on to serve in the US Army. Students accepted at West Point undergo a highly selective process[6] and come from all 50 states and several US territories and foreign militaries. To gain acceptance to West Point, an applicant must secure a nomination from her member of Congress or a senator, pass a physical fitness exam, and show leadership potential. There is no tuition at West Point, and students receive free housing, room, board, and a stipend for book and living expenses. The 47-month experience between reception day and graduation comprises military and leadership training, academic rigor, and physical challenges. Students graduate West Point with a bachelor's in science degree but can major in a number of topics as diverse as English, management, mechanical engineering, or economics.

In addition to academic coursework, students take classes in military science and physical education. They

[4] A flipped classroom is where the students watch a video cover class materials in advance and then ask questions or discuss material in class.

[5] Patterson, Pope and Feudo (2019) use this random assignment of course order across semesters to show that cadets are more likely to select a major in a course that they take first.

[6] The median SAT math score is a 650, which is similar to state flagship institutions or selective liberal arts colleges such as Williams College, Virginia Polytechnic Institute, University of Michigan, or Davidson College.

are required to either be an NCAA recruited athlete or compete in intramural sports and also receive feedback on their leadership and interpersonal skills. West Point determines class rank by a weighted average of performance in academic, military, and physical domains. In exchange for four years of funded college education, students agree to serve as active duty commissioned officers in the US Army for five years and three years on reserve status, beginning at the rank of second lieutenant. West Point graduates serve in various occupational branches within the Army such as quartermaster, infantry, or armor, and their choice of major does not influence jobs available to them, which is generally determined by class rank.

While the flexibility in class load, sequence, and professor is generally greater for students at other colleges, the remote versus online struggle struck colleges across the world at the outset of COVID-19. Students at West Point maintained little control over their academic schedules, were still required to complete 24 required ”core courses” including economics, and continued their mandatory attendance at every class period.[7] These courses are similar to general education classes at civilian colleges and are largely completed in their first two years of the four-year timeline. While these conditions may differ from other colleges, required responses to COVID-19 by West Point leadership (i.e., 50 percent classroom capacities, pivoting courses online, testing and social distancing measures) are similar to what other college administrators faced.

3 Data and Summary Statistics

The data from this experiment come from three sources. The first source is class grades throughout the fall 2020 semester. West Point requires that each course have a unified curriculum with the same graded events across instructors and class sections, and instructors report all grades for each assignment into a central system. We used these grades to show how online education affected multiple types of academic performance including daily homework, problem sets, midterms, and a final exam.

We also received data from the Office of the Dean[8] regarding student demographics and pre-West Point academic achievement. We combined these data to help ensure the internal validity to our experiment and to test for heterogeneous treatment effects. Finally, we administered a post-course survey to the students. This survey was voluntary and students received extra credit for completing it. Their responses were only seen by the researchers, and they provided their identification number so we could link responses to the other two data sets.

Table 1 shows summary statistics for student demographics. First, we standardized all scores of graded events (including the final grade) such that grades have a mean zero and standard deviation of one. Demographically,

[7]The Principles of Economics class consists of concepts from introductory microeconomics, macroeconomics, and personal finance, with all instructors using *Principles of Economics, 8th edition* by Mankiw (2017).

[8]The Dean of the Academic Board at West Point is a one-star general and serves as the chief academic officer at the Academy. This position is similar to a provost.

our sample is representative of the West Point student body, with 23 percent of our sample female, 14 percent black, 3.3 percent Hispanic, and 5.6 percent Asian.

One of the key pillars of a student's education is physical training, and West Point recruits heavily for intercollegiate and club sports; thus 29 percent of our sample are NCAA athletes. Each student must either participate in an intercollegiate, club, or intramural team. West Point also hosts the US Military Academy Preparatory School (USMAPS) that offers community college-like courses for prospective students who may need some additional academic development. These students can then re-apply to the Academy after one year of study. In our sample, 14.2 percent of students were USMAPS graduates. Finally, West Point offers a program for enlisted soldiers with only a high school diploma who are younger than 24 and still single. The program provides the opportunity to attend West Point, earn a bachelor's degree, and commission as an officer. Of the students in our sample, 16.9 percent were prior enlisted.

Finally, it is helpful to understand the academic ability of students before coming to West Point. West Point draws a geographically diverse student body because each member of Congress or senator can only nominate a fixed number of constituents to the military academies. To compare students from across the country, West Point derived an index called the College Entrance Examination Rank (or CEER score). The CEER score is a composite of the SAT, ACT, and high school GPA and has a maximum score of 800.[9] The average CEER score in our sample is a 623.

4 Experimental Setting and Random Assignment

Like many colleges and universities during the COVID-19 pandemic, West Point limited classroom capacity to incorporate social distancing measures. However, unlike other institutions, it still needed to carry out its unique mission to provide physical and military training, and thus students returned to campus for the fall 2020 semester. In addition to the students, around one-half of the faculty are military faculty members who West Point selects from the regular Army, attend graduate school, and then teach at the Academy for two to three years. These officers generally teach the required introductory classes and are vital for student mentorship (Carrell, Page and West, 2010; Jones and Kofoed, 2020; Mansour et al., Forthcoming).West Point's faculty also include permanent military and civilian professors with PhDs who do teach the introductory courses but focus their efforts on upper division courses.

As previously mentioned, West Point students have little control over their daily academic schedules and this policy did not change during the COVID-19 pandemic. We received permission to use this already existing random assignment to assign students to either an in-person or online class section. In addition, to allow for in-person instruction, each instructor agreed to teach their four-section teaching load[10] half online and half in person. This

[9]Hanser and Oguz (2015) show that the CEER score is actually highly predictive of a prospective cadet's probability of graduation and academic success but not necessarily of early promotion as an Army officer.

[10]Four instructors taught only two sections, one online and one in person. Two instructors taught three sections, with two online and one in

agreement allowed for two enhancements to our experimental setting. First, it prevented our faculty from teaching eight class sections while allowing them to provide some in-person instruction. Second, it created counterfactual classrooms where one could see a given instructor teach both in person and online.[11] This setting allows us to control for instructor talent, experience, grading acumen, and familiarity with the course material. We can thus control for instructor intangibles and isolate the treatment effect of being in an online course by including instructor fixed effects.

Given the classroom constraints, each in-person and online class allowed for 12 and 18 students, respectively. Thus of our sample of 551 students, 337 were included in an online section, while 214 were in an in-person section. Most class hours were balanced with respect to the number of online or in-person sections except for the early morning (7:55AM) class hour, which offered more in-person than online sections. However, for every class hour that the economics department offered the introductory course, both teaching modalities were present and the Academy randomly assigned students between the formats. In total, we randomly assigned 551 students across 12 instructors for a total of 36 classrooms.

In addition to the overall summary statistics described in the previous section, Table 1 also compares students assigned to online or in-person class sections. Column (2) shows means and standard deviations for demographic characteristics for students in online sections, Column (3) shows similar summary statistics for students in in-person sections, and Column (4) shows differences and whether they are statistically significant. The first row shows standardized final grades defined as total points earned over a possible 1,000 . We find that students in online class sections earned a grade 0.234 standard deviations lower than those in in-person classes (around 1.638 percentage points), or around a half of a +/– grade. This result is statistically significant in a differences in means test. Given our randomization, this effect is potentially causal, but we test for robustness by adding a host of fixed effects and exogenous controls.

The remaining rows show balance across student demographics. We show that online and in-person classes are balanced by gender and race. Differences between teaching modalities are small (around 1 percentage point for female, Hispanic, and Asian students and one-tenth of 1 percent for black students) and statistically insignificant. Regarding other student characteristics, we find that NCAA athletes are 6 percentage points more likely to be assigned an in-person section because they are assigned early morning classes to afford training in the afternoon and there were more in-person sections than online sections in the early morning hours. This difference, however, is not statistically significant. We find that former prep school students are 0.002 percentage points more likely to be in an online class, prior enlisted students are 2.2 percentage points more likely to be in an in-person class, and the average CEER score is 1.861 points (from a mean of 622) more in online classes.

Next, we show covariate balance and test the efficacy of the experiment by regressing whether a student was

person.

[11]There were two faculty members, however, who either taught all of their classes online or all in person. We dropped students enrolled in these classes from our experiment for a cleaner estimate. However, our results are robust when we include these students in the sample.

in an online class on a number of covariates, instructor fixed effects, class day fixed effects, and time of day fixed effects. Table 2 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 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.

5 Empirical Approach

Given the random assignment of students to either a class section taught online or in person, the econometric model is straightforward. To estimate the causal effect of taking the Principles of Economics class online instead of in person, we estimate the following regression model:

$$y_{ijdt} = \beta_0 + \beta_1 \text{online}_{ijdt} + \gamma_j + \xi_d + \phi_t + X_{ijdt} + \varepsilon_{ijdt}, \tag{1}$$

where y_{ijdt} is the score that student i earns on a graded event (midterm, homework, final exam, final grade) from instructor j in a class taught on day d at class time t. β_1 is a parameter that estimates the causal effect of taking the class online as opposed to in person. We include γ_j or instructor fixed effects. These instructor fixed effects will control for instructor attributes that are invariant across teaching modalities including teaching experience, familiarity with the material, or personality. Given West Point's alternating day one or day two class schedules, we include class day fixed effects (ξ_d) to control for different exam versions and instructor "learning by doing." Next, we include time of day fixed effects (ϕ_t) to control for differences in NCAA athlete class time assignment and any student performance differentials attributed to when they take the class. Finally, we include a vector (X_{ijdt}) of exogenous controls and student demographics such as gender, race, NCAA athlete, prior military service, and preparatory school graduate.

Our paper is unique in that instructors taught half of their sections online and the other half in person. This assignment allowed similar students to see a given instructor in both modalities. Thus the instructor fixed effects simulate an experiment where a hypothetically identical student would learn from an instructor in person and then compare that student to a counterfactual who only saw the instructor online. Since the curriculum is uniform and graded events are identical in content and delivery, the only difference in the experiment would be the mode of instruction. One exception is that instructors used different exam versions across class days, requiring us to adjust our estimates using day fixed effects. One limitation to this method is that instructors could have either taught differently in an online environment or were more uncomfortable online than they were in person. Thus the treatment effect could be capturing both changes in instruction inherent in an online environment and differences in instructor practices between the modalities. While our experiment cannot disentangle these two effects, it is representative of the experiences of many instructors across higher education during the pandemic.

For inference, we would normally adjust our standard errors for correlation between students in a given classroom by clustering at the classroom level. However, we find that the clustered standard errors are smaller than the robust standard errors, a sign of a small cluster problem. To address this concern, we follow the findings of Cameron, Gelbach and Miller (2008) and estimate wild bootstrapped standard errors along with robust standard errors. We report p-values for inference with both the robust standard errors and wild bootstrapping in each of our results tables for the treatment effect of being in an online course. For transparency, Table A1 includes our main results table using classroom-level clustered standard errors.

6 Results

In this section, we will review the results from our experiment. First, we present our main results from the overall sample. Next, we show results by academic ability (looking at CEER percentile and a quantile regression). Finally, we explore our main result as it pertains to different types of student demograhics.

6.1 Main Results

Table 3 presents the main results. In Column (1), we show that online instruction reduced a student's final grade by 0.236 standard deviations or around 1.650 percentage points (out of 100). This result corresponds to about one-half of a +/– grade. Next, to control for differences in instructor talent, attentiveness, or experience, we add instructor fixed effects to our model. This addition reduces the estimated treatment effect to –0.220 standard deviations, a slight decrease in magnitude.

In the next columns, we add additional fixed effects to control for each level of random assignment. West Point generally randomly assigns students to both the day and hour that they will attend a specific class. First, we add a class day fixed effect that controls for the day of the week and the version of an exam that a student took. This fixed effect is also helpful because most of the instructors are new military faculty who usually teach two sections on one day and two the next day. These fixed effects control for any "learning by doing" for a specific lesson. This specification shows an increased learning gap of 0.223 standard deviations. Next, we control for a class time of day fixed effect that controls for any imbalance across class hour such as West Point's preference to assign NCAA athletes a morning class. This specification decreases the learning gap by 0.005 standard deviations to 0.218 standard deviations or 1.526 percentage points or around a half of a +/– grade. Finally, we add a host of exogenous controls including whether a student is female, black, Hispanic, Asian, or an NCAA athlete; attended the USMAPS; or was previously enlisted in the US Army. This specification, included in Column (5), shows a learning gap of 0.215 standard deviations or 1.505 percentage points.

These results show gaps in learning and academic achievement between online and in-person instruction and are robust to the addition of instructor fixed effects, class day fixed effects, time of day fixed effects, and exogenous controls. This robustness shows the effectiveness of our experiment and the corresponding balance across class sections and instruction modality. These results are substantial and in line with other research that examines the effects of technology in the classroom in an experimental setting (such as Carter, Greenberg and Walker, 2017; Figlio, Rush and Yin, 2013).

6.2 Results by Graded Events

We next examine results for each type of graded event. These results are helpful because they can help us understand if the type of assignments may be more conducive for online learning and can give us a sense of whether students adapt to the online learning environment and improve their grades. Figure 1 visually shows the effect of online learning for each type of graded event: problem sets, daily homework, all exams, and for each exam individually. We display each point estimate along with its corresponding 95 percent confidence interval.

Table 4 shows point estimates and standard errors for each regression result with a different type of graded event as an outcome. Column (2) shows the combined scores from six problem sets. These problem sets tested a student's knowledge at the end of each curriculum block. Students were free to work with each other and use resources such as the textbook, but they had to turn in their own work and report which classmates they asked for help. We find that students in online classroom scored 0.106 standard deviations lower than their peers in in-person classes. Column (3) shows results for daily homework assignments. These assignments were different from problem sets since students completed them using multiple attempts with hints and suggestions. The daily homework was worth five points a day and is usually viewed as a measure of student engagement and persistence as instructors graded them for completion as opposed to accuracy. The learning software grades homework automatically, so there is no instructor discretion. We find that online instruction lowered a student's homework grade by 0.211 standard deviations, a result similar to the overall course grade.

Next, we examine the effects for exams. During the course, instructors distributed exams via online learning software regardless of teaching modality. Exams were identical and all students took exams online in their dorm room. There were two versions of the final exam. However, students do not take the final exam with their classmates and instead West Point randomly assigns exam days and times with no regard to instructor or teaching modality. All students took the final online in the instructional software provided by the textbook publisher. For the final exam, we added exam version fixed effects.

In Columns (5) through (8), we find negative and statistically significant effects in performance (with the third midterm and the final as exceptions) on these higher-stakes exams, while the gap between online and in-person classes narrows over time. While the confidence intervals for individual exams do overlap with each other, the point estimates suggest an upward trend in learning for online students. This result still shows a gap in learning from online instruction, but that gap may lessen once students gain experience with the online framework, develop new study habits, or perhaps seek help from students in an in-person course. However, the grade reduction on these high-stake exams was large enough to significantly reduce a student's final grade.

6.3 Results by Academic Ability

Our main result shows that overall there is a large reduction in student performance in online classes. However, we are interested to see if this result is different for those of lower academic acheivement (as measured by the CEER score that West Point uses for admission). We estimate out model by CEER quantiles and using a quantile regression.

6.3.1 Subsample Analysis by Prior Academic Ability

Next, we consider how the online environment differently affected students of varying academic ability. We estimate the same empirical model but with different subsamples based on a student's CEER score. In our sample, the median CEER score was a 630 out of 800 and the 25th percentile was a 568. We estimate our model for cadets below the 25th and 50th and then above the 50th percentile to understand which part of the academic distribution may be driving our results.

Table 5 displays the results for our effect by CEER score. We use both subsample analysis and specifications with an interaction of the main effect with whether a student lies within a given CEER quantile to estimates differences between groups. Column (2) shows that learning gaps are greater for those students whose high school academic preparation was in the bottom quarter of the distribution. Here, we find that being in an online class section reduced their final grades by 0.267 standard deviations, translating to around 1.869 percentage points of the student's final grade. This result has large implications for a student's GPA, particularly at West Point where a student's occupation and first posting within the Army is generally determined by their GPA. Next, in Column (3) we interact whether a student is in the bottom quartile of the CEER distribution with the indicator for whether a student took the class online. The interaction term has a coefficient of –0.232 but is not statistically significant because of our small sample size. Column (4) shows a similar result for students whose CEER score was in the bottom half of the distribution. Here, we find that online learning reduced a student's final grade by 0.269 standard deviation. Column (5) also shows that the coefficient associated with interaction term is –0.203. This result shows that the lower half of the academic distribution is less than the upper half but is not statistically significant.

Finally, in Column (6) we find, however, that among students whose CEER scores were in the top half of the distribution, there is a smaller and statistically insignificant effect of online learning. We estimate that for these students, online instruction reduced their final grades by 0.149 standard deviations or about 1.043 percentage points, an effect that is indistinguishable from zero. In Column (7), we interact whether a student is in the top half of the CEER distribution with whether the class was taught online. The interaction term is positive but statistically insignificant. These results imply that the pivot to online education may have had larger effects for academically at-risk students. However, these findings must be weighed against the threat of infection from COVID-19 and the possibility of transmission to areas and vulnerable populations around a college campus. To ensure that these students were not harmed grade-wise and career-wise, we adjusted the grades of those students who were assigned an online class section upward according to these point estimates.

6.3.2 Analysis by Grade Distribution

Next, we consider how online education affected students based on their performance in the course. We first present these results visually in Figure 2. In this figure, we plot a cumulative distribution function by whether a student was in an online or in-person class section. In the figure, the dotted line represents online class sections, while the solid line represents in-person sections. We find that at every point on the distribution (except for a few points on the very bottom), the in-person sections dominate the online sections. In the graph, the mass of the online distribution is located below the mass on the in-person grades. We find that the gap is the greatest at the bottom half of the grade distribution and closes at the top.

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 6 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.

6.4 Heterogeneous Effects by Student Demographics

Next, we consider how online learning affects various demographic groups differently. One advantage of West Point data is the rich demographic information available on students. We chose several subsamples that allow us to understand the differences among students and confirm previous findings about at-risk students. Table 7 displays the results for each demographic subsample. First, we test differences across student gender and display these findings in Columns (2) and (3). We find that online instruction is more detrimental to male students than female students. We find that female students in an online class earn a grade that is 0.051 standard deviations lower than female students in an in-person section, a result that is statistically insignificant. For males, we find that online instruction reduced their grades by 0.266 standard deviations or 1.862 percentage points, a statistically significant result.

Next, we examine our results by race. Columns (4) and (5) show results for black and white cadets. We find a slight increase of 0.005 standard deviations for black students in remote sections, a result that is statistically insignificant. However, we do find that online classes reduced white students' final grades by 0.277 standard deviations or 1.939 percentage points. This difference in results could be that we are underpowered regarding black students since there are only 80 in our sample.

The next demographic characteristics are somewhat unique to West Point but may shed some additional insights on our results, particularly the evidence that our result is a product of lower academic ability. First is whether a student is an NCAA athlete. Given some slight covariate balance discussed early, we want to ensure that NCAA athletes do not drive our results. Column (7) shows the results for NCAA athletes where we estimate this reduction as 0.226 standard deviations, a result very similar to our main result. When we only consider non-NCAA athletes in Column (8), we estimate a similar effect of 0.212 standard deviations. These results are statistically insignificant from each other and the baseline results, thus allaying the concerns about covariate imbalance.

Finally, we consider students who previously enlisted in the Army (and thus postponed enrolling in college for up to five years) and those who first attended the preparatory school (equivalent to a year of community college). Column (9) displays the results for prior enlisted students. We find that online instruction reduced their final grades by 0.448 or 3.316 percentage points. Column (10) shows that online instruction reduced final grades for preparatory school attendees by 0.378 standard deviations (or around 2.646 percentage points) or roughly an entire +/– grade. Since both of these groups are among those that are the most academically at risk, this result provides further evidence that online instruction has a larger effect on lower ability students.

7 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. Second, we use a post-course survey to measure student responses to questions on preparing for class, concentration, and connectedness to instructors and peers.

7.1 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.[12] 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 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.

Table 8 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.

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

7.2 Mechanisms: Post-Course Survey

Finally, we distributed a post-course survey to better understand whether student study habits and their perceptions about the instructor and their peers changed in the online environment. This survey was not mandatory, but instructors did give their students bonus points for completing it. Four hundred and two students completed the evaluation for a response rate of 72.95 percent. There was also coverage for all class sections and instructors. Of those students who completed, 60.45 percent were online compared to 61.16 percent in the full sample. Regarding the observable characteristics, survey respondents were slightly more likely to be female (25.37 versus 22.88 percent) and less likely to be black (12.19 versus 14.41 percent) but were otherwise balanced against the whole sample. Thus while we do not have perfect compliance, the subsample should be representative of the whole.

Table 9 shows the effects of online instruction on a host of outcomes that we hope to help explain why student performance decreased. First, we ask students for an estimated number of minutes that they needed to prepare for each class meeting. Column (1) shows that students in online classes reported studying 2.352 more minutes per day, a result that is statistically insignificant. However, it is interesting that students spent slightly more time studying and yet still received a lower grade.

We next asked students to rank their ability to concentrate in class. Students could respond with a "1" for not being able to concentrate at all up to a "5" if they were perfectly able to concentrate. The average student responded with a "3.515" with a standard deviation of 0.951. We estimate both an OLS regression and an ordered probit to estimate the effect of online instruction on concentration. Columns (2) and (3) show that online instruction lowered student perception of being able to concentrate by 0.557 and 0.692 in the OLS and ordered probit, respectively. These results are roughly three-fourths of a standard deviation decline that is statistically significant.

On our survey, we also explore how online instruction affected the student-instructor relationship. We asked students to rank how connected they felt to their instructor from 1 to 5. On average, students ranked how connected they felt as 3.798 with a standard deviation of 0.831. Columns (4) and (5) show results for these estimates. We find that students rated their connection with instructors lower by 0.342 and 0.500 (OLS and ordered probit, respectively) on the five-point scale. This result translates to around two-thirds of a standard deviation and is statistically significant.

We then asked students whether they felt that their instructor cared about them. Again, students could respond with a 1 if they felt that their instructor did not care about them at all up to a 5 if they strongly felt their instructor cared. The average student response was a 4.2 with a standard deviation of 0.784. Columns (4) and (5) show these results for both the OLS and ordered probit. We find that being in an online class reduced whether a student felt that their instructor cared about them. We also find that students' perception about their instructors caring about them declines by 0.126 or 0.200 points using the OLS or ordered probit, respectively, a result that is marginally insignificant. This result amounts to a decrease of about a quarter of a standard deviation.

Finally, we asked students to rank their connection to their peers with a similar scale as before. The average student response was a 2.699 with a standard deviation of 1.039. Columns (6) and (7) show the results for the OLS and ordered probit models, respectively. We find that the online environment reduced how connected a student felt toward their peers by 0.500 or 0.534 points, or a decrease of roughly a half a standard deviation or 3.5 percentage points.[13]

Results from the post-course survey are interesting because the instructor fixed effects are controlling for instructor attributes, teaching styles, and personalities while comparing student perception and connectivity across teaching modalities. These results show that one cost of online education during the pandemic was student satisfaction, concentration, and many of the intangible benefits a professor provides in an in-person class. Many researchers have shown that instructor quality, characteristics, and intensity of interaction are important for student success, particularly for lower income and minority students (Vlieger, Jacob and Stange, 2020; Figlio and Schapiro, 2021; Fairlie, Hoffmann and Oreopoulos, 2014; Bettinger and Long, 2005; Carrell and West, 2010). We find that online instruction during the COVID-19 pandemic may have disrupted these peer and instructor interactions.

8 Conclusion

The global pandemic caused widespread disruption on all levels of education. The pivot from in-person to online instruction was unplanned and forced students and instructors into an uncertain, online environment. This change disrupted both K-12 and university education, increased educational inequality (Bacher-Hicks, Goodman and Mulhern, 2021), caused political pressure to re-open (Collier et al., 2021), and displaced parental labor supply (especially among working mothers) (Zamarro and Prados, 2021). However, randomized control trials that can estimate differences in student learning between the two modalities causally are rare. West Point's unique institutional mission and makeup allowed us the opportunity to randomize students into either an online or in-person teaching environment. In our experimental setting, West Point's registrar randomly assigned students into class sections and our faculty agreed to split their teaching loads between the two modalities. This feature allowed us to compare students with the same instructor in a course with uniform lesson plans, graded events, and exams, the only difference being whether a class is online or in person. This setting is advantageous because we did not rely on volunteers or an instrumental variable, and our results are not specific to a single instructor.

This paper makes several findings regarding the efficacy of online learning during the COVID-19 pandemic. We find that online learning reduced final grades by a half +/– grade, a result that increases for lower ability students.

[13]Hardt, Nagler and Rincke (2020) show that randomized peer tutoring was helpful during the COVID-19 pandemic. This type of intentional peer support networks could be helpful to overcome the negative peer connectivity effects in this paper.

We also find evidence that the result is concentrated among students most academically at risk. We use questions from a post-course survey to show that online students experienced reduced concentration; they also felt less connected to their instructors and peers and claimed their instructors cared less about them. These results show the limitations of online learning, especially with the little time instructors had to prepare and adjust teaching styles and pedagogy. From an ethical perspective, we should note that while it is Academy-wide policy to randomly assign students to classes, we did adjust the final grade of students in online sections according to our findings and prioritized lower CEER score students for in-person classes during the 2021 spring semester.

There are some limitations in our West Point setting. First, our student and instructor sample may not be representative of the universe of higher education. West Point students tend to have higher SAT/ACT scores, do not need to worry about employment, and have free room and board. The pivot to online education (and the accompanied recession caused by the pandemic) may have increased college student unemployment and student financial need (Gurantz and Wiegla, Forthcoming; Clelan and Kofoed, 2017). These extra stressors and insecurities may have caused students to not complete courses or not return for the next semester. Also, West Point requires students to attend classes regardless of instructional modality, and instructors report attendance to commanding officers that ensure compliance. Finally, while there were some outbreaks of COVID-19, West Point's closed campus environment, close medical attention, and quarantining limited infection, and campus did not close partway during the semester. While these factors may cause a West Point student's experience to differ from the average college student, we argue that our findings may serve as a helpful lower bound of the negative impact of online instruction. Our results may indicate that other college students, particularly those from disadvantaged backgrounds, may have fared worse.

Our results indicate there are limitations to the effectiveness of online education, especially during a global pandemic. Like many college instructors, those at West Point had little time to prepare lessons and develop pedagogy to more effectively teach online. Our findings are agnostic to whether those teaching styles could mitigate the learning loss that we observe. However, college administrators and higher education policymakers should cautiously consider whether increased online offerings are actually in a student's best interest and the most effective way to deliver a quality college education.

References

Alpert, William T., Kenneth A. Couch, and Oskar R. Harmon. 2016. "A Randomized Assessment of Online Learning." *American Economic Review: Papers & Proceedings*, 106(5): 378–382.

Bacher-Hicks, Andrew, Joshua Goodman, and Christine Mulhern. 2021. "Inequality in Household Adaptation to Schooling Shocks: Covid-Induced Online Learning Engagement in Real Time." *Journal of Public Economics*, 193: 104345.

Bettinger, Eric P., and Bridget Terry Long. 2005. "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students." *The American Economic Review: Papers and Proceedings*, 95(2): 152–157.

Bettinger, Eric P., Lindsay Fox, Susanna Loeb, and Eric S. Taylor. 2017. "Virtual Classrooms: How Online College Courses Affect Student Success." *American Economic Review*, 107(9): 2855–2875.

Bilen, Eren, and Alexander Matros. 2021. "Online cheating amid COVID-19." *Journal of Economic Behavior & Organization*, 182: 196–211.

Brady, Ryan R., Michael A. Insler, and Ahmed S. Rahman. 2017. "Bad Company: Understanding Negative Peer Effects in College Achievement." *European Economic Review*, 98: 144–168.

Cacault, M. Paula, Christian Hildebrand, Jeremy Laurent-Lucchetti, and Michele Pellizzari. Forthcoming. "Distance Learning in Higher Education: Evidence from a Randomized Experiment." *Journal of the European Economic Association*, 1–51.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *The Review of Economics and Statistics*, 90(3): 414–427.

Carrell, Scott E., and James E. West. 2010. "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors." *Journal of Political Economy*, 118(3): 409–432.

Carrell, Scott E., Marianne E. Page, and James E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *Quarterly Journal of Economics*, 125(3): 1101–1144.

Carrell, Scott E., Teny Maghakian, and James E. West. 2011. "A's from Zzzz's? The Causal Effect of School Start Time on the Academic Achievement of Adolescents." *American Economic Journal: Economic Poliyc*, 3: 62–81.

Carter, Susan Payne, Kyle Greenberg, and Michael S. Walker. 2017. "The impact of computer usage on academic performance: Evidence from a randomized trial at the United States Military Academy." *Economics of Education Review*, 56: 118–132.

Christensen, Clayton M., and Henry J. Eyring. 2011. *The Innovative University: Changing the DNA of Higher Education from the Inside Out.* San Francisco, California:Jossey-Bass.

Clelan, Elizabeth B., and Michael S. Kofoed. 2017. "The Effect of the Business Cycle on Freshman Financial Aid." *Contemporary Economic Policy*, 35(2): 253–268.

Coates, Dennis, Brad R. Humphreys, John Kane, and Michelle 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.

Collier, Daniel, Dan Fitzpatrick, Madison Dell, Sam Snideman, Christopher Marsicano, and Robert Kelchen. 2021. "We Want You Back: Uncovering the Influences on In-Person Instructional Operations in Fall 2020." *C2i Working Paper Series, No. 210101*, 1–57.

Cowen, Tyler, and Alex Tabarrok. 2014. "The Industrial Organization of Online Education." *American Economic Association, Papers & Proceedings*, 104(5): 519–522.

Crow, Michael M., and William B. Dabars. 2015. *Designing the new American University.* Baltimore, Maryland:Johns Hopkins University Press.

Deming, David J., Claudia Goldin, Lawrence F. Katz, and Noam Yuchtman. 2015. "Can Online Learning Bend the Higher Education Cost Curve?" *American Economic Review: Papers & Proceedings*, 105(5): 496–501.

Deming, David J., Michael Lovenheim, and Richard Patterson. 2020. "The Competitive Effects of Online Education." In *Productivity in Higher Education*. , ed. Caroline M. Hoxby and Kevin Stange. Chicago, Illinois:The University of Chicago Press.

Dutton, John, Marilyn Dutton, and Jo Perry. 2002. "How Do Online Students Differ From Lecture Students?" *Jounral of Asynchronous Learning Networks*, 6(1): 1–20.

Fairlie, Robert W., Florian Hoffmann, and Philip Oreopoulos. 2014. "A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom." *American Economic Review*, 104(8): 2567–2591.

Figlio, David, and Morton Schapiro. 2021. "Staffing the Higher Education Classroom." *Journal of Economic Perspectives*, 35(1): 143–162.

Figlio, David, Mark Rush, and Lu 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.

Gurantz, Oded, and Christopher Wiegla. Forthcoming. "how Have FAFSA Submissions Differed during COVID-19." *Educational Researcher*.

Haggag, Kareem, Richard W. Patterson, Nolan G. Pope, and Aaron Feudo. 2019. “Attribution Bias in Major Decisions: Evidence from the United States Military Academy.” *IZA DP No. 12174*.

Hanser, Lawrence M., and Mustafa Oguz. 2015. “United States Service Academy Admissions: Selecting for Success at the Military Academy/West Point and as an Officer.” *RAND Corporation Publication*.

Hardt, David, Markus Nagler, and Johannes Rincke. 2020. “Can Peer Mentoring Improve Online Teaching Effectiveness? An RCT During the COVID-19 Pandemic.” *Working Paper*.

Jones, Todd R., and Michael S. Kofoed. 2020. “Do Peers Influence Occupational Preferences? Evidence from Randomly-Assigned Peer Groups at West Point.” *Journal of Public Economics*, 184: 1–17.

Joyce, Ted, Sean Crockett, David A. Jaeger, Onur Altindag, and Stephen D. O’Connell. 2015. “Does Classroom Time Matter?” *Economics of Education Review*, 46: 64–77.

Kofoed, Michael S., and Elizabeth mcGovney. 2019. “The Effect of Same-Gender or Same-Race Role Models on Occupation Choice: Evidence from Randomly Assigned Mentors at West Point.” *Journal of Human Resources*, 54(2): 430–467.

Lyle, David S. 2007. “Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point.” *Review of Economics and Statistics*, 89(2): 289–299.

Lyle, David S. 2009. “The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point.” *American Economic Journal: Applied Economics*, 1(4): 69–84.

Mangrum, Daniel, and Paul Niekamp. 2020. “JUE Insight: College Student Travel Contributed to Local COVID-19 Spread.” *Journal of Urban Economics*, 1–34.

Mankiw, N. Gregory. 2017. *Principles of Economics, 7th Edition.* Boston, Massachusetts:Cengage Learning.

Mansour, Hani, Daniel Rees, Bryson Rintala, and Nathan Wozny. Forthcoming. “The Effects of Professor Gender on the Post-Graduation Outcomes of Female Students.” *Industrial and Labor Relations Review*.

Patterson, Richard W., and Robert M. Patterson. 2017. “Computers and productivity: Evidence from laptop use in the college classroom.” *Economics of Education Review*, 57: 66–79.

Patterson, Richard W., Nolan G. Pope, and Aaron Feudo. 2019. “Timing is Everything: Evidence from College Major Decisions.” *IZA DP No. 12069*, 1–64.

Russell, Thomas L. 1999. *The No Significant Difference Phenomenon As Reported in 355 Research Reports, Summaries and Papers.* Raleigh, North Carolina:North Carolina Press.

Setren, Elizabeth, Kyle Greenberg, Oliver Moore, and Michael Yankovich. Forthcoming. "Effects of Flipped Classroom Instruction: Evidence from a Randomized Trial." *Education Finance and Policy*, 1–54.

Vlieger, Pieter De, Brian Jacob, and Kevin Stange. 2020. "Measuring Instructor Effectiveness in Higher Education." In *Productivity in Higher Education*. , ed. Caroline M. Hoxby and Kevin Stange. Chicago, Illinois:The University of Chicago Press.

Zamarro, Gema, and Maria J. Prados. 2021. "Gender Differences in Couples' Division of Childcare, Work, and Mental Health During COVID-19." *Review of Economics of the Household*, 19: 11–40.

Tables and Figures

Table 1: Summary Statistics Comparing Full Sample, Online, and In Person Classrooms

	(1) Full Sample mean/sd	(2) Online mean/sd	(3) In Person mean/sd	(4) Difference b/se
Final Grade (std.)	0.000	-0.092	0.144	0.236***
	(1.000)	(1.003)	(0.980)	[0.087]
Final Grade (%)	83.367	82.726	84.376	1.650***
	(6.998)	(7.018)	(6.860)	[0.608]
Online	0.612	1.000	0.000	
	(0.488)	(0.000)	(0.000)	
Female	0.230	0.226	0.238	0.013
	(0.422)	(0.419)	(0.427)	[0.037]
Black	0.140	0.139	0.140	0.001
	(0.347)	(0.347)	(0.348)	[0.030]
Hispanic	0.033	0.027	0.042	0.015
	(0.178)	(0.161)	(0.201)	[0.016]
Asian	0.056	0.050	0.065	0.015
	(0.231)	(0.219)	(0.248)	[0.020]
NCAA Athlete	0.290	0.267	0.327	0.060
	(0.454)	(0.443)	(0.470)	[0.040]
Prep School	0.142	0.142	0.140	-0.002
	(0.349)	(0.350)	(0.348)	[0.031]
Prior Enlisted	0.169	0.160	0.182	0.022
	(0.375)	(0.367)	(0.387)	[0.033]
CEER Score	623.806	623.083	624.944	1.861
	(73.377)	(70.981)	(77.155)	[6.419]
Observations	551	337	214	551

This tables shows differences in standardized course grades and covariate balance across treatment and control groups. Means are in the first row with standard deviations in parentheses. The last column shows covariate balance across treatment and control with standard errors below in brackets. Statistical significance levels are as follows: * for $p<0.10$, ** for $p<0.05$, and *** for $p<0.01$. We also include a cadet's CEER score as a measure of academic ability. CEER Score is an index combining measures of academic potential such as high school GPA, high school class rank, SAT score, and ACT score.

Table 2: 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
Female	-0.017							0.003	0.007	0.009	0.002
	(0.050)							(0.051)	(0.051)	(0.051)	(0.050)
Black		-0.001						-0.011	-0.011	-0.015	-0.006
		(0.060)						(0.063)	(0.063)	(0.063)	(0.062)
Hispanic			-0.115					-0.120	-0.110	-0.105	-0.153
			(0.120)					(0.119)	(0.125)	(0.123)	(0.124)
Asian				-0.067				-0.083	-0.070	-0.071	-0.026
				(0.092)				(0.094)	(0.099)	(0.098)	(0.096)
NCAA Athlete					-0.069			-0.080^{*}	-0.074	-0.083^{*}	-0.060
					(0.046)			(0.048)	(0.049)	(0.049)	(0.048)
Prep School						0.004		0.157	0.175^{*}	0.169	0.176
						(0.060)		(0.109)	(0.106)	(0.102)	(0.107)
Prior Enlisted							-0.037	-0.148	-0.162^{*}	-0.156^{*}	-0.163
							(0.056)	(0.101)	(0.097)	(0.093)	(0.100)
Instructor FE	No	No	No	No	No	No	No	No	Yes	Yes	Yes
Class Day FEs	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	Yes
F-Stat	0.119	0.001	0.926	0.530	2.238	0.005	0.441	0.886	0.843	0.909	0.812
F-Stat, p-value	0.730	0.981	0.336	0.467	0.135	0.941	0.507	0.517	0.552	0.499	0.578

Robust 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 3: Main Effects for Online Instruction

	(1) Final Grade	(2) Final Grade	(3) Final Grade	(4) Final Grade	(5) Final Grade
Online	-0.236***	-0.220**	-0.223***	-0.218**	-0.215**
	(0.086)	(0.087)	(0.086)	(0.089)	(0.084)
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
Robust SEs p-values	0.007	0.011	0.010	0.014	0.011
Wild Bootstrapped SEs p-values	0.007	0.010	0.009	0.012	0.013

Robust 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 cadet 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.

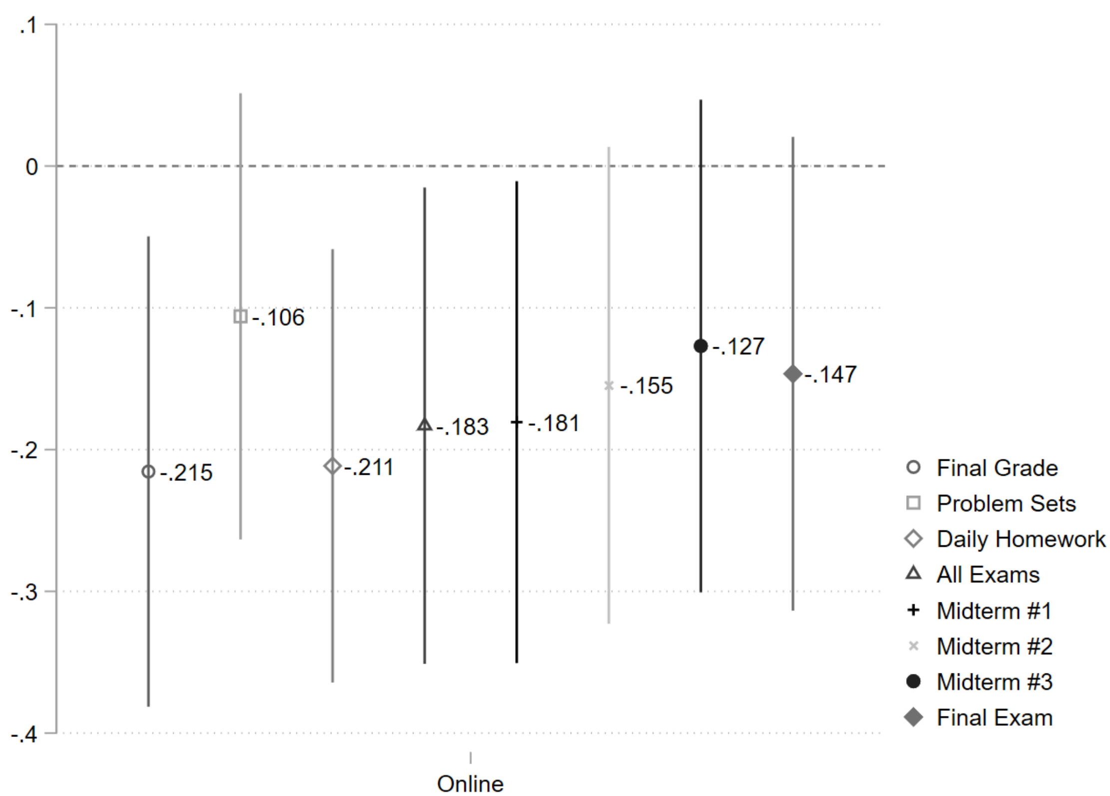
Table 4: Effects of Online Instruction on Each Graded Event

	(1) Final Grade	(2) Problem Sets	(3) Daily Homework	(4) Total Exams	(5) Midterm #1	(6) Midterm #2	(7) Midterm #3	(8) Final Exam
Online	-0.215**	-0.106	-0.211***	-0.181**	-0.181**	-0.155*	-0.127	-0.147*
	(0.084)	(0.080)	(0.078)	(0.087)	(0.087)	(0.086)	(0.088)	(0.085)
Instructor FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class Day FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Time of Day FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exog. Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	551	551	551	551	551	551	551	551
R^2	0.173	0.139	0.183	0.132	0.142	0.108	0.109	0.152
Robust SEs p-values	0.011	0.186	0.007	0.037	0.037	0.071	0.152	0.086
Wild Bootstrapped SEs p-values	0.013	0.204	0.002	0.040	0.038	0.077	0.154	0.087

Robust 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 for each graded event. Exogenous controls include whether a cadet 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.

Figure 1: Effects of Online Instruction on Each Graded Event



This figure shows the result for the effect of online learning on college student academic achievement for each graded event. Each point estimate represents a coefficient estimate for a regression that contains instructor fixed effects, class day fixed effects, and exogenous controls. Exogenous controls include whether a cadet is female, black, Hispanic, Asian, or an NCAA athlete; attended the US Military Academy Preparatory School; or was prior enlisted. Confidence levels come from robust standard errors.

Table 5: Effects for Online Instruction by Academic Ability

	(1) Overall	(2) CEER<25th	(3) CEER<25th	(4) CEER<50th	(5) CEER<50th	(6) CEER>50th	(7) CEER>50th
Online	-0.215**	-0.267*	-0.187*	-0.269**	-0.105	-0.149	-0.293***
	(0.084)	(0.155)	(0.098)	(0.112)	(0.108)	(0.112)	(0.108)
CEER $<$ 25th percentile			-0.511***				
			(0.144)				
Online $\times$ CEER $<$ 25th percentile			-0.232				
			(0.174)				
CEER $<$ 50th percentile					-0.741***		
					(0.125)		
Online $\times$ CEER $<$ 50th percentile					-0.203		
					(0.150)		
CEER $>$ 50th percentile							0.746***
							(0.125)
Online $\times$ CEER $>$ 50th percentile							0.181
							(0.150)
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
Observations	551	132	551	270	551	277	551
R^2	0.173	0.360	0.226	0.193	0.314	0.088	0.310
Robust SEs p-values	0.011	0.088	0.057	0.017	0.331	0.183	0.007
Wild Bootstrapped SEs p-values	0.013	0.090	0.053	0.017	0.351	0.193	0.012

Robust 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 for whether a student was below the lowest academic ability quartile or the median score or above the median as measured by the CEER score. CEER Score is a measure of academic potential that combines high school GPA, high school class rank, SAT score, and ACT score. Exogenous controls include whether a cadet 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.

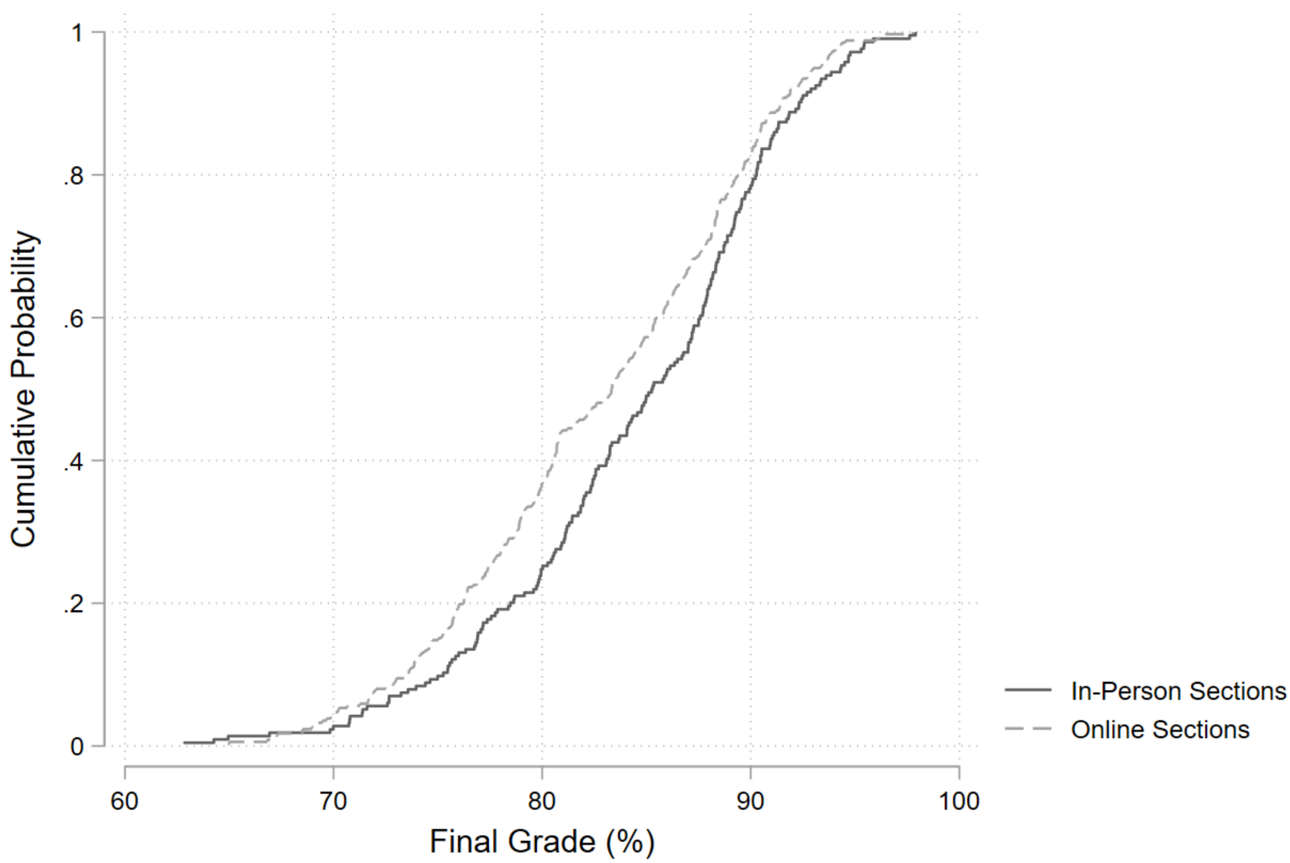
Table 6: Quantile Regression for Final Course Grade

	(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.

Figure 2: Cumulative Distributive Function for Final Course Grade by Teaching Modality



This figure shows cumulative distribution functions for a student's final course grade by teaching modality. The solid line represents grades from an in-person class section, while the dotted line represents online class sections. The figure shows that the bulk of the online mass is among lower final grades and the in-person distribution dominates at all levels.

Table 7: Heterogeneous Treatment Effects for Online Education

	(1) Overall	(2) Female	(3) Male	(4) Black	(5) White	(6) NCAA Athlete	(7) Non-NCAA Athlete	(8) Prior Service	(9) Non-Prior Service	(10) Prep School
Online	-0.215**	-0.090	-0.266***	0.005	-0.277***	-0.226**	-0.212***	-0.448**	-0.157***	-0.378*
	(0.084)	(0.179)	(0.075)	(0.188)	(0.067)	(0.107)	(0.065)	(0.191)	(0.057)	(0.206)
Instructor FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class Day FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time of Day FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exog. Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	551	127	424	77	403	160	391	93	458	78
R^2	0.173	0.292	0.175	0.230	0.110	0.248	0.174	0.384	0.127	0.443
Robust SEs p-values	0.011	0.618	0.001	0.979	0.000	0.041	0.002	0.022	0.009	0.073
Wild Bootstrapped SEs p-values	0.013	0.625	0.007	0.983	0.006	0.109	0.008	0.029	0.030	0.073

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table shows the heterogeneous treatment effect for online learning on college student academic achievement for various types of students. 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 8: 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 9: Mechanisms: Effects of Online Instruction on Post-Course Survey

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Study Time (mins)	Student Concentration		Connected to Instructor		Instructor Care		Connected to Peers	
	OLS	OLS	Ordered Probit	OLS	Ordered Probit	OLS	Ordered Probit	OLS	Ordered Probit
Online	2.352	-0.557***	-0.692***	-0.342***	-0.500***	-0.126	-0.200	-0.500***	-0.534***
	(1.699)	(0.100)	(0.124)	(0.084)	(0.121)	(0.084)	(0.127)	(0.119)	(0.129)
Instructor FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class Day FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time of Day FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exog. Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	402	402	402	402	402	402	402	402	402
R^2	0.104	0.146	0.079	0.179	0.079	0.100	0.046	0.123	0.045
Robust SEs p-values	0.167	0.000	0.000	0.000	0.000	0.134	0.116	0.000	0.000
Wild Bootstrapped SEs p-values	0.161	0.000	0.000	0.000	0.000	0.121	0.121	0.001	0.000

Robust 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 how students felt about their course experience from a post-course survey. Study Time (mins) is how many minutes a student prepared for each class meeting. Student Concentration is how students rated their in-class concentration on a five-point scale. Connected to Instructor is how students rated their connection to their instructor on a five-point scale. Instructor Cares represents how much students felt their instructor cared about them measured on a five-point scale. Connected to Peers is how students rated their connection to peers on a five-point scale. 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. Class Day FEs control for which day the class meet.

Table A1: Main Effects for Online Instruction: Clustered Standard Errors

	(1) Final Grade	(2) Final Grade	(3) Final Grade	(4) Final Grade	(5) 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.