2018

Who Succeeds in Distance Learning? Evidence from Quantile Panel Data Estimation

Bacolod, Marigee; Mehay, Stephen; Pema, Elda

Bacolod, Marigee, Stephen Mehay, and Elda Pema. "Who succeeds in distance learning? Evidence from quantile panel data estimation." Southern Economic Journal 84.4 (2018): 1129-1145.
http://hdl.handle.net/10945/59938

Southern Economic Journal 2018, 84(4), 1129–1145
DOI: 10.1002/soej.12264

Who Succeeds in Distance Learning? Evidence from Quantile Panel Data Estimation

Marigee Bacolod,* Stephen Mehay,† and Elda Pema‡

This study provides a comprehensive analysis of the distributional effects of distance learning (DL) on academic success, as measured by course grades and completion. Using data of over 1.2 million courses taken by about 200,000 U.S. Navy sailor-students at more than 1800 U.S. institutions during 1994–2007, we find that distance delivery of education is associated with poorer outcomes. At the mean, DL delivery is associated with 0.19 lower course grade points; however, the mean effect masks the more pronounced negative effects of DL in the bottom two-thirds of the distribution—where DL lowers grades by as much as 0.8 points. Using variation only among marginal students—those who tend to fail some of the courses that they take, our estimates indicate traditional face-to-face delivery is associated with 2.4 times greater likelihood of successful course completion than if it was delivered distant. These findings suggest that targeting DL courses to certain students may be more cost-effective.

JEL Classification: I20, I23

1. Introduction

The push for increased use of distance education in colleges and universities has grown rapidly since the mid-1990s, particularly with recent enthusiasm for massive open online courses (MOOCS).[1] Distance Learning (DL) is a mode of delivering instruction to students who are not physically present in a traditional classroom setting.[2] In 2014, about 28% of all postsecondary students took some of their courses online; of these, 14% took all of their courses online, while another 14.5% took at least one course online (Allen and Seaman 2016; U.S. Department of Education 2016).

The appeal of DL programs lies in its promise for increased access to higher education and the more efficient utilization of existing facilities (Deming et al. 2015). For students employed full-

* Graduate School of Business and Public Policy, Naval Postgraduate School, 555 Dyer Road, Monterey, CA 93943, USA; E-mail: mbacolod@nps.edu corresponding author.

† Graduate School of Business and Public Policy, Naval Postgraduate School, 555 Dyer Road, Monterey, CA 93943, USA; E-mail: smehay@nps.edu.

‡ Graduate School of Business and Public Policy, Naval Postgraduate School, 555 Dyer Road, Monterey, CA 93943, USA; E-mail: epema@nps.edu.

Received July 2017; accepted February 2018.

[1] For instance, the New York Times dubbed 2012 "The Year of the MOOC." http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html?smid=pl-share.

[2] The U.S. Department of Education defines distance education as "education that uses one or more technologies to deliver instruction to students who are separated from the instructor and to support regular and substantive interaction between the students and the instructor synchronously or asynchronously." Online provision via MOOCS is a recent development of a type of distance learning.

time, such as those used in this study, a major benefit of DL is the time flexibility it allows. On the other hand, these benefits may come at the cost of lower student learning, either for DL students overall and/or for some subgroups of students. Indeed, the growing literature finds no robust consensus on the effects of distance education. Some studies find poorer performance for online students (see, e.g., Coates et al. 2004; Xu and Jaggars 2011, 2013), while others find few or no significant differences between the two instructional modes (Bowen et al. 2014; Figlio, Rush, and Yin 2013). Meanwhile, several studies including those that find no overall difference in the effect of distance education, report sizeable differences in the effects of DL across various subgroups of students (defined by race, gender, or ability level), suggesting substantial heterogeneity in the impacts of distance delivery of education (Bettinger et al. 2017; Figlio, Rush, and Yin 2013; Xu and Jaggars 2014).

In this study, we explicitly account for individual heterogeneity and provide a comprehensive analysis of the distributional effects of DL on student success, as measured by course grades and course completion rates. Because it is unlikely that DL delivery of higher education has a uniform effect throughout the conditional student performance distribution, we estimate quantile effects on student grades, in addition to average effects using panel data.

We utilize course-level data on over 1.2 million courses taken by about 200,000 U.S. Navy enlisted personnel at more than 1800 U.S. higher education institutions between 1994 and 2007. Our sample of enlistees—who are in the data because they voluntarily enrolled in off-duty college classes through the Navy's Tuition Assistance (TA) program—is particularly well-suited for estimating the distributional impact of DL. Unlike civilian samples where populations of DL and resident students are highly self-selected and face very different time constraints, in our sample individual sailors take both DL and face-to-face courses during the same period while working similar jobs and hours. Navy enlistees are free to choose both the courses and institutions they attend, similar to the U.S. college-going population. Their participation in DL is necessitated by largely random job rotation or deployment.[3] Most importantly from an estimation standpoint, we can exploit variation in course performance from many sailors who take multiple courses (some DL and some traditional) during the same year, in the same subject, and while employed in the same job position. Nonetheless, we view our estimates as more suggestive rather than causal effects of the impact of distance learning.

In previous studies with panel data that employ quantile regression techniques, the individual fixed effects typically enter as an additive term. The additive fixed effect, however, has the interpretation that the quantile treatment effects do not vary with that fixed effect. In other words, the potentially heterogeneous effects of DL across the student performance distribution is not allowed to vary with student ability—a fairly restrictive assumption. Our empirical strategy does not impose this condition of additive separability of the individual fixed effect. Instead it utilizes the within-group variation for identification as proposed by Powell (2016) to flexibly condition on individual, subject, and year fixed effects. Thus, our empirical strategy and data allow us to take into account individual heterogeneity in estimating how DL delivery affects outcomes across the student performance distribution.

---

[3] Job rotations (called sea-shore rotations) are prescribed by a sailor's career path in an occupational specialty or rating, for example, 48 months at sea followed by 36 at shore for electricians. There is, however, huge variation in the timing of these rotations due to a myriad of factors, such as ship repair schedules, whether the rotation involves U.S. or overseas assignment, and changing military needs. This variation generates the randomness in the timing of actual job rotations that affects the choice in DL versus face-to-face enrollment.

Our estimates confirm prior findings of the negative association of DL delivery on student success at the mean. In addition, we find that the negative effect of DL is particularly being driven by the bottom half of the student ability distribution. At the mean, DL delivery is associated with 0.19 lower course grade points; however, the mean effect masks the more pronounced negative effects of DL in the bottom tail of up to 0.8 lower grade points. When restricting the sample to marginal students—those who tend to fail some of the courses that they take, our estimates indicate that traditional face-to-face delivery is associated with 2.4 times greater likelihood of successful course completion than DL delivery of the same course. These findings suggest that targeting DL courses to certain individuals may be more cost-effective without sacrificing outcomes.

The rest of the article is organized as follows. Section 2 provides a brief literature review. Section 3 describes the Tuition Assistance program and our data. Section 4 lays out our empirical specifications, and section 5 presents our results. Section 6 concludes.

2. Prior Research

Although numerous prior studies have attempted to assess the effectiveness of online education on student learning outcomes, few of the early studies applied rigorous statistical techniques in estimating the effects of distance education. A meta-analysis by the U.S. Department of Education concluded that of the 1000 studies they reviewed, only 50 used credible randomization or quasi-randomization strategies to address the differential selection of students into online versus traditional courses (Means et al. 2010). A key difficulty in estimating the impact of distance education is that the typical student taking an online course is often very different from the typical student taking a face-to-face course. DL students tend to be older, more mature, have families, and are full or part-time employed. Self-selection into DL, ability differences, and differing time constraints may account for the lower academic performance of DL students, not the course delivery method itself. Thus, causal implications in the early literature are problematic due to weaknesses in methodology, especially the failure to control for student heterogeneity (Jaggars and Bailey 2010).

More recent studies, using both observational and experimental data have adopted research designs to address the selection issues. Anstine and Skidmore (2005) apply instrumental variables techniques to deal with endogeneity in enrollment decisions and find that test scores are lower in online sections than in face-to-face sections of the same economics and statistics classes. Coates et al. (2004) apply an endogenous switching model to compare test scores for students in online and face-to-face sections of an economics course. They find the average student performs worse in online classes. They also find that the type of student most likely to select online classes performs better in that instructional setting, suggesting positive selection.

Xu and Jaggars (2011) use multilevel propensity score matching techniques to analyze academic outcomes in the first math or English courses for Virginia community college students. The study finds that online students are significantly less likely to complete the courses and, among course completers, have lower grades. In a second study, Xu and Jaggars (2013) analyze community college students in Washington State and account for student heterogeneity using fixed effects techniques for school, term, and subject. In addition, they apply instrumental variables techniques to deal with potential bias from student selection of the instructional mode for a single course. They find that online students have lower persistence and course grades. They also find that self-selection exerts a downward bias on performance estimates in unadjusted models.

More recent studies have used experimental designs to estimate the learning effects of online courses. Bowen et al. (2014) randomly assigned students at six public universities into online and face-to-face sections of a statistics course. The online course was a hybrid that included interactive learning materials supplemented by a one hour per week face-to-face session. They find no differences in mean outcomes (course pass rates, final exam scores, and scores on a standardized statistics test) between the two formats, with the exception of higher course completion rates for hybrid students.[4]

Figlio, Rush, and Yin (2013) randomly assigned students in an introductory microeconomics course at a selective university, either to a section using live lectures or to a section offering online videos of the same lectures. Although they find no significant differences in overall student achievement between the two instructional modes, they report significant negative effects of the online format among Hispanics, males, and lower-achieving students (those with low prior GPAs).

Bettinger et al. (2017) analyze a large database of students enrolled in a major for-profit college using an instrumental variables strategy to address student selection. The study also controls for all other aspects of the course (class size, syllabi, textbooks, and professor assignment), which often are unobserved in prior studies, and estimates fixed effects for course, term, major, and home campus. The study finds that online students have lower course grades in the online course as well as in follow-on courses, and have lower persistence in college. The study also finds greater variability of course grades in the online courses and that the negative effects of online courses are larger for students with lower prior GPAs.

In one of the few studies to focus specifically on the distributional effects of online courses, Xu and Jaggars (2014) estimate individual and course fixed effects models using data on community college students. They find that overall student performance is weaker in online courses, and that these performance gaps are larger for males, younger students, Blacks, and students with lower prior GPAs.

Although both observational and experimental studies have found that certain subgroups perform worse in DL courses, these distributional effects have not been the focus of prior research. Against this backdrop, our study thus contributes to the literature by assessing the overall effects of online courses on student outcomes, and also by focusing specifically on the distributional effects of distance education among students.

Our study offers several advantages over prior research. First, the data consists of employees of a single large organization with a rigid internal labor market (all hiring at the entry point, no lateral entry, and promotions from within). All employees work full time and take employer-subsidized college courses while off-duty. A second feature of the data is the large scale of the TA program—our sample consists of over 200,000 TA program participants who enrolled in about 1.2 million courses in over 1800 colleges during the study period. Third, the sample population consisting of those who take DL and non-DL courses, are relatively homogeneous in terms of age, work experience, and demographic background. In terms of demographics and academic preparedness, Navy TA students are most similar to civilian community college students, with the major exception of a higher proportion of males than females. Fourth, data are available to control

[4] Cosgrove and Olitsky (2015) compare outcomes for Principles of Economics students in three instruction modes: face-to-face, blended classes (which integrate face-to-face and online components), and web-enhanced. None of the sections were fully online. Using scores on tests administered before, during, and at the end of the courses, they find students in traditional classes have greater knowledge retention and achieve greater test score gains compared to blended students.

for work demands and job responsibilities of students, allowing us to explicitly account for their job skills, and time constraints. We also have an explicit measure of student ability in the Armed Forces Qualifying Test (AFQT) that all Navy recruits are required to take. This is an improvement over much of the literature that uses SAT/ACT scores or prior GPAs. Together with our quantile regression panel data framework that we discuss in detail below, our data and strategy allows for estimation of the heterogeneous effects of DL across the student ability distribution.

3. Background and Data

Our data includes all Navy enlisted personnel who ever took off-duty college classes via the Navy's Tuition Assistance (TA) program during 1994–2007. The TA program is the largest employer-funded college education program in the United States. Each year over 340,000 active duty personnel participate in the Defense Department's Voluntary Education program (VOLED). The largest component of VOLED is the TA program, which reimburses service members for the costs of college classes. Annually, about 18% of all Navy enlistees enroll in college classes via the TA program.

The TA program has seen rapid growth and several important policy changes during the period under study. In 2002, the Navy increased tuition reimbursement rates from 75 to 100%. Between 2000 and 2007, enrollment in undergraduate college courses supported by TA grew by 50%. By 2008, 52,481 Navy enlisted personnel enrolled in 152,698 classes with assistance from the TA program.

Meanwhile, similar to the civilian higher education market, there has been rapid growth in the number of distance learning courses taken by TA students. Between 2000 and 2007, the number of DL classes taken by sailors via TA grew ten-fold whereas the number of traditional classes fell by 29%. In 2006, for the first time, enrollment in college classes taught via DL exceeded enrollment in traditional classroom settings. The growth in distance education was stimulated by Navy policies aimed at improving access to postsecondary education, particularly by increasing the availability of online courses. In 1999, the Navy initiated formal partnerships with colleges to provide online courses and degrees. Called the Navy College Program Distance Learning Partnership (NCPDLP), this program expanded rapidly in the 2000s; by 2007 it included 34 colleges.[5]

Sailors are free to take courses and complete degrees from any postsecondary institution.[6] Our data covers all 1,960,592 course enrollments between 1994 and 2007 via the Navy's TA program. The sample is then restricted to enlisted personnel taking undergraduate college courses, leaving 1,641,740 course-level observations. These courses are taken at 1885 different postsecondary institutions. The Naval Education and Training Command provided course and student-level data on all TA participants and the Defense Manpower Data Center provided demographic data on all TA users.

As seen in Table 1, only 27% of all courses in our data were taken via distance learning. Of those sailors who have taken at least 1 DL course, 67% of all their courses are taken via DL.

---

[5] NCPDLP schools agree to offer A.A. or B.A. degrees online and to link degree programs with Navy enlisted occupations (called ratings).

[6] All colleges serving military students belong to the Service Members Opportunity College program. All member institutions agree to accept transfer credits from other member schools and to restrict residency requirements for degrees.

Table 1. Descriptive Statistics by DL Status

Panel A. At the Individual Level

	Sailor-Students Who Never Took DL		Sailor-Students With At Least 1 DL		All Sailor-Students	
Variable	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Percentage Courses taken via DL	0	0	0.669	0.344	0.271	0.395
Percentage Complete Course	0.817	0.325	0.797	0.318	0.809	0.322
Female	0.22	0.414	0.241	0.428	0.228	0.42
Black	0.324	0.468	0.24	0.427	0.29	0.454
Hispanic	0.121	0.326	0.132	0.338	0.125	0.331
AFQT score	61.462	19.604	62.833	19.354	62.04	19.511
Missing AFQT	0.152	0.359	0.093	0.29	0.128	0.334
Percentage Bottom Tercile on AFQT	0.055	0.228	0.051	0.221	0.054	0.225
Percentage Middle Tercile on AFQT	0.438	0.496	0.447	0.497	0.441	0.497
Percentage Top Tercile on AFQT	0.355	0.478	0.409	0.492	0.377	0.485

Summary statistics are for 216,346 enlisted American sailors who participated in the Navy's Tuition Assistance Program between 1994 and 2007.

Panel B. At the Individual-Course Level (n = 1, 296, 287)

	Traditional Courses		DL Course		All Courses	
Variable	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
GPA	3.215	0.977	3.037	1.161	3.175	1.024
Percentage Complete Course	0.893	0.309	0.825	0.38	0.878	0.328
GPA, conditional on completion	3.324	0.791	3.264	0.841	3.311	0.803
Female	0.238	0.426	0.256	0.436	0.242	0.428
Black	0.334	0.472	0.24	0.427	0.312	0.463
Hispanic	0.12	0.325	0.124	0.33	0.121	0.326
Age at Enrollment	11.666	15.059	12.069	15.598	11.759	15.185
Missing Age	0.606	0.489	0.608	0.488	0.607	0.488
AFQT score	61.881	19.658	63.26	19.388	62.216	19.602
Missing AFQT	0.155	0.362	0.092	0.29	0.141	0.348
Percentage Bottom Tercile on AFQT	0.053	0.225	0.05	0.218	0.053	0.223
Percentage Middle Tercile on AFQT	0.427	0.495	0.437	0.496	0.429	0.495
Percentage Top Tercile on AFQT	0.364	0.481	0.42	0.494	0.377	0.485
Business Course	0.138	0.345	0.167	0.373	0.145	0.352
STEM Course	0.353	0.478	0.343	0.475	0.35	0.477
Humanities Course	0.384	0.486	0.399	0.49	0.388	0.487
Rank = E4	0.194	0.395	0.149	0.356	0.183	0.387
Rank = E5	0.305	0.461	0.303	0.46	0.305	0.46
Rank = E6	0.223	0.416	0.267	0.442	0.233	0.423
Rank = E7	0.11	0.313	0.146	0.353	0.118	0.323
Rank = E8	0.03	0.17	0.043	0.202	0.033	0.178
Rank = E9	0.009	0.094	0.013	0.115	0.01	0.1
FY 1994	0.008	0.091	3.35e–06	0.002	0.006	0.08
FY 1995	0.093	0.291	9.38e–05	0.01	0.072	0.258
FY 1996	0.087	0.282	2.14e–04	0.015	0.067	0.25
FY 1997	0.083	0.276	0.004	0.066	0.065	0.246
FY 1998	0.088	0.283	0.012	0.109	0.07	0.256
FY 1999	0.091	0.287	0.02	0.138	0.074	0.262
FY 2000	0.09	0.286	0.027	0.163	0.075	0.264
FY 2001	0.089	0.284	0.042	0.201	0.078	0.268
FY 2002	0.089	0.285	0.07	0.256	0.085	0.279
FY 2003	0.073	0.26	0.101	0.301	0.079	0.27

Table 1. (*Continued*)

Panel B. At the Individual-Course Level (n = 1, 296, 287)

Variable	Traditional Courses		DL Course		All Courses	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
FY 2004	0.069	0.253	0.155	0.362	0.088	0.284
FY 2005	0.059	0.236	0.195	0.396	0.091	0.287
FY 2006	0.026	0.158	0.103	0.304	0.043	0.204
FY 2007	0.056	0.23	0.271	0.444	0.105	0.307

Summary statistics are for courses taken by 216,346 enlisted American sailors who participated in the Navy's Tuition Assistance Program between 1994 and 2007.

In Table 1, Panel B shows that successful completion rates in traditional courses is greater (89%) than in DL courses (82%). We categorize a course as not successfully completed if the outcome is a drop, withdrawal, or a grade of incomplete, nonpassing, or an "F."

Course grades are also on average lower in DL courses compared to traditional courses, whether or not the course was completed. Figure 1 further illustrates the gap in these course grades by mode of delivery. A significantly greater proportion of As are earned by students in traditional resident classes, compared to distance classes. The gap in density narrows at Bs, narrows even further to a negligible difference at Cs, and reverses at Ds and Fs. In other words, a greater proportion of resident courses result in A grades, while a greater proportion of DL courses result in Ds and Fs.

In terms of demographics, Blacks are less likely to take DL courses while there does not appear to be a significant difference in mode of education delivery for Hispanics. Examining the AFQT distribution, a greater proportion of those in the top tercile take DL classes, suggesting that simply conditioning on AFQT as a measure of ability will not be sufficient to distinguish heterogeneous effects of DL across the student performance distribution.

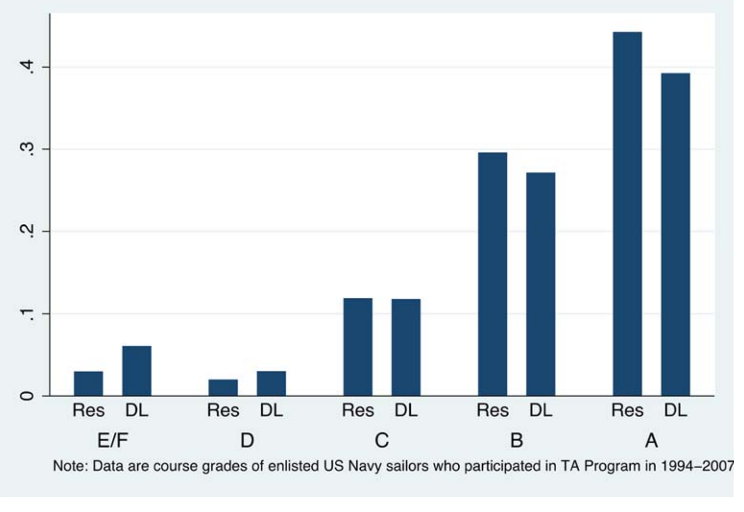

Figure 1. Distribution of Course Grades by DL versus Resident. [Color figure can be viewed at wileyonlinelibrary.com]

4. Empirical Specifications

To estimate the relationship between distance learning and successful course completion, we estimate fixed-effects logit models. The specification we estimate is:

$$P\left(complete_{ij}=1\right)=F\left(\alpha_i+\beta_1\ D_{ij}+\beta_2\ X_{ij}\right), \tag{1}$$

where $complete_{ij} = 1$, if student i successfully completes course j and 0 otherwise, $D_{ij} = 1$ if the mode of delivery of course j is distance education and 0 otherwise, and X_{ij} are individual and course-specific characteristics such as gender, rank (pay grade), and age at the time of the course. We also include year fixed effects to control for changes over time in program policies and other factors. For example, the Global War on Terror, which began in 2003, increased the operational tempo and deployments of sailors. In specification (Eqn. 1), our main parameter of interest is β_1.

To estimate the effect of DL across the distribution of student course grades, we employ the quantile regression panel data (QRPD) estimator. The quantile regression coefficient is simply the distributional analog of ordinary least squares (OLS), which is the difference in mean outcomes between DL and resident courses conditional on other regressors. In unconditional quantile regression, if $F(y)$ is the cumulative distribution function (CDF) of the outcome Y, then the τth quantile of $F(y)$ is the smallest y_τ value such that $F(y_\tau) \geq \tau$. The unconditional quantile regression coefficient on DL is then the difference between the τth quantiles of the DL course outcomes' CDF F_1 and the resident outcomes' CDF F_0:

$$F_1^{-1}(\tau)-F_0^{-1}(\tau)=y_{\tau 1}-y_{\tau 0}.$$

Quantile regressions are particularly useful for estimating differences in outcomes in a way that allows for heterogeneity across the entire student performance distribution. The most common way prior research has addressed distributional concerns is to estimate mean impacts for different subgroups, for example, by race or by prior GPA, as described earlier. However, these mean impacts does not allow for the heterogeneity to flexibly vary across the distribution. That is, mean subgroup analyses may miss nuances in estimating in which subgroup positive or negative effects of DL are particularly concentrated.

Meanwhile, the inclusion of fixed effects in the context of quantile regressions has long posed a challenge. Most previous studies simply assume this fixed effect is additive, and use the residual from a regression of the outcome variable on the fixed effect as the quantile regressions' dependent variable. However, the main reason one is interested in estimating quantile regressions is to identify impacts at points other than the mean. Specifying an additive fixed effect undermines this intent as it separates the error term into different components, and the quantile parameters are then constrained to vary based only on the non-fixed part of the disturbance term.

In terms of an econometric specification, the quantile function assuming an additive fixed effect is

$$Y_{ij}=\alpha_i+\beta_1(\tau)D_{ij}+\beta_2(\tau)X_{ij}. \tag{2}$$

In contrast, the structural quantile function of interest is

$$Y_{ij}=\alpha_i(\tau)+\beta_1(\tau)D_{ij}+\beta_2(\tau)X_{ij}. \tag{3}$$

In other words, the effect of unobserved student ability on outcomes is allowed to vary across the distribution. Quantile estimators with additive fixed effects estimate the distribution of

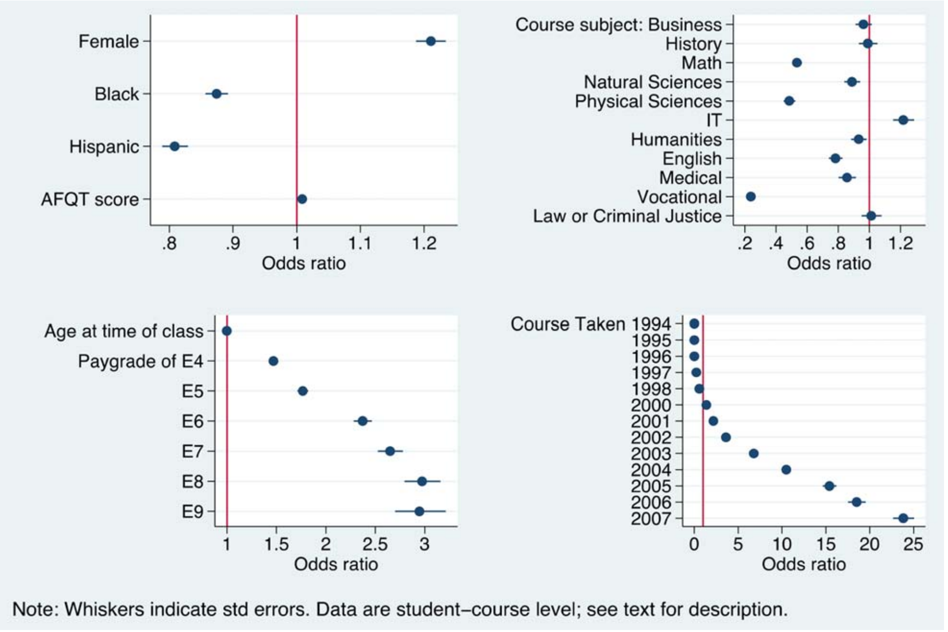

Figure 2. Logit Odds Ratios of Probability of DL Course. [Color figure can be viewed at wileyonlinelibrary.com]

$(Y_{ij} - \alpha_i)|D_{ij}$ instead of $Y_{ij}\,|D_{ij}$. However, students with grades at the top of the $(Y_{ij} - \alpha_i)|D_{ij}$ distribution may instead be at the bottom of the $Y_{ij}|D_{ij}$ distribution. Given our research question, we are interested in the effect of DL across the student performance distribution, $Y_{ij}|D_{ij}$.

On the other hand, estimation of Equation 3 is non-trivial. Fortunately, Powell (2016) shows that exploiting the panel structure can substantially reduce the number of parameters that need to be estimated using Generalized Method of Moments (GMM). Our strategy is to thus flexibly condition on individual fixed effects for identification, taking advantage of the variation from students who take both DL and resident courses. We employ the Nelder-Mead optimization procedure to form estimates of $\beta_1(\tau)$.[7] For comparison we also estimate Equation 2 and in practice include individual, year, and subject fixed effects.

In addition, per Firpo (2007), with selection on observables we can also obtain efficient estimates of the quantile coefficients by weighting with the inverse of the propensity scores. To obtain these scores, we first estimate a sailor's probability of taking a course via DL using a logit, with course subject, age, rank, fiscal year, and the demographics in Table 1 as controls. These logit estimates are illustrated in Figure 2, showing the growth in DL course-taking between 2000 and 2007. In the quantile regressions, we then weight observations taken via DL with the inverse of the predicted probability from the logit. Courses taken via resident learning are weighted by one over one minus the propensity score.

Finally, we note that ideally we would like to be able to measure student learning as outcomes, of which course grades are but noisy measures. Moreover, the assignment of grades tend to vary considerably from program to program, instructor to instructor, and importantly for our

[7] The authors thank David Powell for kindly sharing the STATA *.ado file for this estimation.

context, distance versus face-to-face even for the same instructor teaching the same course. We are unable to directly control for these factors as our data do not have information on instructors and college major. However, individual fixed effects can mitigate some of these concerns to the extent that students self-select into courses across instructors and mode of learning based on their unobservable characteristics. Also, we control for subject as an aggregate proxy for course major.

5. Results

DL and Course Completion

In Table 2, we report our estimates of Equation 1 relating DL delivery of courses with the probability a student-sailor successfully completes that course. The first column of Table 2 assumes F(.) is linear and estimates Equation 1 via OLS. Holding all other factors constant, distance delivery of a course is associated with 0.078 lower probability of course completion, relative to face-to-face delivery of that course. Note that this estimate holds constant course subject, fiscal year of enrollment, and a sailor's job skill level and supervisory responsibilities (Rank). Given how the course-level data are distributed across subjects, in our estimation we aggregated subjects into four groups: Business, STEM (math, physical and natural sciences, medical, IT), Humanities

Table 2. Effect of DL on Course Completion

	(1)	(2)	(3)
Variables	Fixed Effects (OLS)	Fixed Effects Logit	FE-Logit Odds Ratios
DL Course	−0.0783***	−0.8740***	0.4173***
	[0.0010]	[0.0129]	[0.0054]
Business Course	−0.0115***	−0.1996***	0.8191***
	[0.0011]	[0.0179]	[0.0147]
STEM Course	−0.0314***	−0.4798***	0.6189***
	[0.0010]	[0.0153]	[0.0095]
Humanities Course	−0.0131***	−0.2399***	0.7867***
	[0.0010]	[0.0151]	[0.0119]
Rank = E4	0.0113***	0.1590***	1.1723***
	[0.0015]	[0.0188]	[0.0220]
Rank = E5	0.0289***	0.4008***	1.4930***
	[0.0020]	[0.0250]	[0.0374]
Rank = E6	0.0455***	0.6192***	1.8575***
	[0.0027]	[0.0368]	[0.0684]
Rank = E7	0.0755***	0.9598***	2.6113***
	[0.0038]	[0.0530]	[0.1384]
Rank = E8	0.0936***	1.1050***	3.0193***
	[0.0051]	[0.0778]	[0.2350]
Rank = E9	0.1166***	1.3301***	3.7814***
	[0.0080]	[0.1289]	[0.4874]
Age at Enrollment	0	0.0002	1.0002
	[0.0003]	[0.0039]	[0.0039]
Observations	1,296,287	560,697	560,697
R^2	0.444	0.0223	0.0223
No. of students	216,346	63,363	63,363

All models include missing indicators and dummies for fiscal years 1994–2007 with 1999 as baseline, and in OLS, a constant. *** indicates significance at 1%, ** at 5%, * at 10%. Standard errors in brackets.

(including history and English), and all other subject fields (e.g., Law or Criminal Justice in Figure 2). The "Other" subjects is the baseline category in all the regression analyses.

In column 2 of Table 2, we present estimates from a fixed effects logit model (with odds ratios presented in column 3). The odds of a student completing a course when it is taken via DL versus resident are 0.4 to 1. Inverting this ratio, the estimated effect of DL in column 3 indicates that a student who takes a course via a traditional classroom is 2.4 times more likely to complete the course than if it was delivered via DL. The larger magnitude of this estimate is not surprising, given that fixed effects logit estimation draws information only from the subset of individuals who complete some of their courses and fail to complete others. That is, identification of fixed effects logit requires variation in both the dependent variable for each student as well and in the independent variables. In contrast, the fixed effects OLS estimation in the first column includes (i) individuals who complete all courses that they take, (ii) those who fail to complete all courses that they take, and (iii) those who complete some of the courses that they take and fail to complete others. Because the fixed effects logit only includes the third subsample, the sample in columns 2 and 3 are less than half of column 1. The very large effect estimated from fixed effects logit indicates that distance delivery of education can have particularly strong negative effects on course completion for marginal students—those that tend to fail some of the courses that they enroll in.

Sailors are least likely to complete courses—whether taken via DL or face-to-face—in STEM, followed by Humanities, Business, and then all other fields. The higher the sailor's Rank at the time of enrollment, the more likely he or she is to complete the course, even conditional on age and individual fixed effects. This is unsurprising given that Rank also proxies for job skill.

DL and Course Grades

Similar to prior research, we find that distance learning is associated with lower grades. First, for contrast we estimate Equation 2, which assumes a linear and additive fixed effect in estimating the quantile function. Figure 3 illustrates the quantile regression coefficients and their standard errors. Figure 3 clearly shows that the bottom end of the conditional student performance

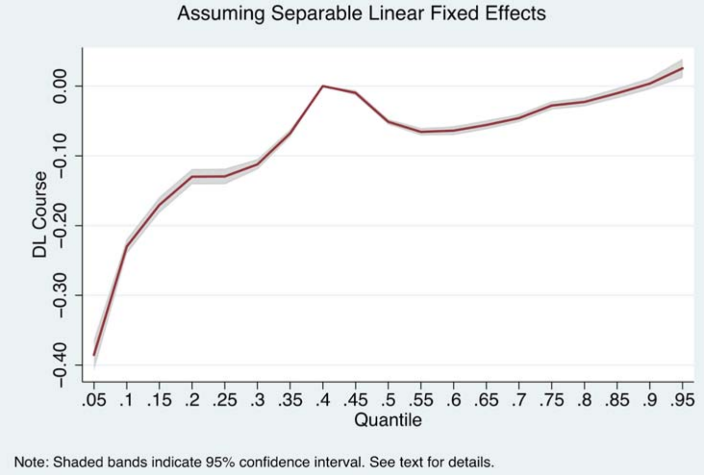

Figure 3. Quantile Regression DL Coefficients. [Color figure can be viewed at wileyonlinelibrary.com]

Table 3. OLS versus QRPD Estimates of the Effect of DL on Course GPA

Panel A. Effect of DL on Course GPA, Whether or Not Completed

Quantile	Mean	5	10	15	20	25	30	35	40	45
DL	−0.1929***	−0.5000***	—0.5744***	−0.5000***	−0.5	−0.1875	−0.8125***	−0.5000***	−0.5000***	−0.1875***
	[0.0038]	[0.0862]	[0.0214]	[0.0151]	[0.3478]	[0.3175]	[0.0017]	[0.0013]	[0.0018]	[0.0011]
Quantile	50	55	60	65	70	75	80	85	90	95
DL	−0.1875***	−0.1875***	−0.1875***	−0.1525***	−0.1508*	0.0001	0	0	0	0
	[0.0020]	[0.0039]	[0.0024]	[0.0064]	[0.0834]	[0.0010]	[0.0000]	[0.0000]	[0.0000]	[0.0000]

Robust standard errors in brackets. *** indicates significance at 1%, ** at 5%, * at 10%. OLS/mean and Quantile Regression with Panel Data (QRPD) models estimated with individual, year, and subject fixed effects. QRPD models estimated using Nelder-Mead optimization method. All models include 1,165,325 sailor-course observations for 200,502 U.S. Navy sailors.

Panel B. Effect of DL on Course GPA, Conditional on Course Completion

Quantile	Mean	5	10	15	20	25	30	35	40	45
DL	−0.0899***	−0.5000***	−0.5000***	−0.5000***	−0.1875***	−0.1875***	−0.5000***	−0.1875***	−0.3125***	−0.5000***
	[0.0031]	[0.0029]	[0.0183]	[0.0067]	[0.0023]	[0.0017]	[0.0017]	[0.0014]	[0.0007]	[0.0007]
Quantile	50	55	60	65	70	75	80	85	90	95
DL	−0.5000***	−0.1875	−0.1417***	−0.1022***	−0.0014	−0.0014	−0.0014	0	0	0
	[0.0014]	[0.1842]	[0.0464]	[0.0188]	[0.0011]	[0.0011]	[0.0011]	[0.0008]	[0.0007]	[0.0000]

Robust standard errors in brackets. *** indicates significance at 1%, ** at 5%, * at 10%. OLS/mean and Quantile Regression with Panel Data (QRPD) models estimated with both individual, year, and subject fixed effects. QRPD models estimated using Nelder-Mead optimization method. All models include 1,117,547 sailor-course observations for 192,293 U.S. Navy sailors.

distribution earns significantly lower grades than those at the top end when they take distance learning courses. However, because these quantile estimates do not allow the fixed effects to vary across the student performance distribution, we focus our discussion on our QRPD estimates.

Table 3 presents our QRPD estimates of the effect of DL on student course grades. In Panel A, we report mean and quantile regression coefficient estimates on course grades regardless of whether or not that course was completed. Panel B reports these estimates conditional on course completion. Both panels include individual, year, and subject fixed effects.

The main takeaway from these estimates is that distance education has a significant negative effect on student course grades on the bottom two-thirds of the course grade distribution. At the mean, DL delivery is associated with 0.19 lower course grade, which translates to one-fifth lower letter grade. However, the negative mean effect masks the substantial difference across the grade performance distribution. In particular, the negative effect of DL at the mean is being driven by the bottom half of the student performance distribution. Examining the estimates across the columns of Table 3, we can see that the negative grade effect of DL is more pronounced for quantiles below the median, with a grade difference of up to 0.8 (at the 30th quantile) lower for DL courses. This lower performance in DL courses translates to a difference of four-fifths of a full letter grade. In contrast, at the upper tail of student performance distribution, there is no statistically detectable difference in course grades by the mode of education delivery. The most able students—up to about the upper 25%, fare just as well if they take classes via distance learning or resident.

Table 4. Heterogeneous Effects of DL on Course Completion

	(1)	(2)	(3)
Group	Fixed Effects (OLS)	Fixed Effects Logit	FE-Logit Odds Ratios
Male	−0.0742***	−0.8708***	0.4182***
	[0.0012]	[0.0157]	[0.0066]
Female	−0.0871***	−0.8510***	0.4269***
	[0.0021]	[0.0222]	[0.0095]
Business Course	−0.0684***	−1.0111***	0.3639***
	[0.0026]	[0.0494]	[0.0180]
STEM Course	−0.0655***	−0.6478***	0.5225***
	[0.0021]	[0.0241]	[0.0126]
Humanities	−0.0876***	−0.9090***	0.4028***
	[0.0018]	[0.0221]	[0.0089]
Bottom Tercile on AFQT	−0.0911***	−0.9875***	0.3723***
	[0.0045]	[0.0543]	[0.0202]
Middle Tercile on AFQT	−0.0799***	−0.8580***	0.4234***
	[0.0016]	[0.0187]	[0.0079]
Top Tercile on AFQT	−0.0788***	−0.8989***	0.4071***
	[0.0016]	[0.0213]	[0.0087]
Hispanic	−0.0883***	−0.8917***	0.4089***
	[0.0030]	[0.0333]	[0.0136]
Black	−0.0984***	−0.9688***	0.3794***
	[0.0020]	[0.0227]	[0.0086]
White	−0.0594***	−0.7473***	0.4735***
	[0.0014]	[0.0204]	[0.0096]

Each cell is a different regression. All models include rank, age at enrollment, missing age, and dummies for fiscal years 1994–2007 with 1999 as baseline, and in OLS, a constant. *** indicates significance at 1%, ** at 5%, * at 10%. Standard errors are in brackets.

Table 5. Heterogeneous Effects of DL on Course Grade

Quantile	Mean	5	10	15	20	25	30	35	40	45
Male	−0.1818***	−0.5000***	−0.875	−0.1875***	−0.2500**	−0.8750***	−0.5000***	−0.1875***	0	−0.5000***
	[0.0045]	[0.0740]	[0.8853]	[0.0193]	[0.1202]	[0.0028]	[0.0370]	[0.0024]	[0.0009]	[0.0009]
Females	−0.2239***	−1.0000***	−0.5000***	−0.5000***	−0.1875**	−0.1875***	−0.1875***	−0.8125***	−0.5000***	−0.5000***
	[0.0074]	[0.0059]	[0.0231]	[0.0759]	[0.0848]	[0.0056]	[0.0045]	[0.0046]	[0.0047]	[0.0026]
Business	−0.2069***	−0.5	−0.5000***	−0.5000***	−0.1875***	−0.5000***	−0.8125***	−0.5000***	−0.1875***	−0.5000***
	[0.0121]	[0.7245]	[0.0190]	[0.0106]	[0.0089]	[0.0068]	[0.0071]	[0.0075]	[0.0050]	[0.0044]
STEM	−0.1103***	−0.5000***	−0.375	−0.1875	−0.5000***	−1.0000***	−0.1875***	0.1875***	−0.1875***	−0.1875***
	[0.0084]	[0.0927]	[0.0000]	[0.1220]	[0.1928]	[0.0025]	[0.0038]	[0.0046]	[0.0035]	[0.0021]
Humanities	−0.2303***	−1.0000***	−0.8125***	−0.5000***	−0.5000***	−0.1875***	−0.5000***	0	−0.3125***	−0.5000***
	[0.0071]	[0.0050]	[0.0766]	[0.0053]	[0.0062]	[0.0040]	[0.0027]	[0.0013]	[0.0027]	[0.0021]
Bottom AFQT Tercile	−0.1877***	−0.8125**	−0.5000***	−0.5000***	−0.5000***	−0.1875***	−0.1875***	−0.5000***	0	−0.1875***
	[0.0111]	[0.3373]	[0.0494]	[0.0103]	[0.0139]	[0.0091]	[0.0068]	[0.0095]	[0.0031]	[0.0048]
Mid AFQT Tercile	−0.1909***	−1.0000***	−0.875	−0.5000***	−0.1875***	−0.5000***	−0.1875***	−0.8125***	−0.5000***	−0.1875***
	[0.0059]	[0.0050]	[1.7096]	[0.0048]	[0.0447]	[0.0054]	[0.0037]	[0.0035]	[0.0034]	[0.0028]
Top AFQT Tercile	−0.1962***	−0.25	−0.5	−0.5000***	−0.8125***	−0.8125***	−0.5000***	−0.3125***	−0.1875***	n.a.
	[0.0060]	[0.5022]	[0.3042]	[0.0044]	[0.0051]	[0.0033]	[0.0025]	[0.0030]	[0.0030]	

Each cell is a separate regression showing the effect of DL for the group indicated in the row header. Robust standard errors are in brackets. *** indicates significance at 1%, ** at 5%, * at 10%. OLS/mean and Quantile Regression with Panel Data (QRPD) models estimated with both individual and year fixed effects. QRPD models estimated using Nelder-Mead optimization method. "n.a." means convergence could not be achieved.

Quantile	50	55	60	65	70	75	80	85	90	95
Male	−0.5000***	−0.1875***			n.a.					
	[0.0068]	[0.0179]								
Females	−0.1875***	−0.1875***	−0.5000***	−0.5000***	−0.5000***	−0.5000***			n.a.	
	[0.0037]	[0.0056]	[0.0065]	[0.0041]	[0.0045]	[0.0053]				
Business	−0.1875***	−0.5000***	−0.5	−0.5	0.8125***		n.a.			
	[0.0092]	[0.0127]	[1.0132]	[0.7719]	[0.0211]					
STEM	−0.5000***	−0.5000***	−0.1875			n.a.				
	[0.0022]	[0.0082]	[0.2731]							
Humanities	−0.6250***	−0.3750***	−1.0000***			n.a.				
	[0.0019]	[0.0948]	[0.0012]							
Bottom AFQT Tercile	−0.1875***	−0.1875***	−0.5000***	−0.5000***	−0.5000***		n.a.			
	[0.0147]	[0.0089]	[0.0065]	[0.0055]	[0.1721]					
Mid AFQT Tercile	−0.1875***	−0.5000***	−0.5000***	−0.5000***			n.a.			
	[0.0025]	[0.0024]	[0.0038]	[0.0031]						
Top AFQT Tercile				n.a.						

Each cell is a separate regression showing the effect of DL for the group indicated in the row header. Robust standard errors are in brackets. *** indicates significance at 1%, ** at 5%, * at 10%. OLS/mean and Quantile Regression with Panel Data (QRPD) models estimated with both individual and year fixed effects. QRPD models estimated using Nelder-Mead optimization method. "n.a." means convergence could not be achieved.

Heterogeneity in Effects of DL

Clearly, our estimates show that focusing just on mean effects misses the substantial heterogeneity in the impact of distance education across the distribution of student outcomes. In addition, Table 4 indicates some heterogeneity in the effect of DL on course completion. As expected, students at the bottom AFQT tercile do significantly worse (in terms of course completion) in DL than students with higher AFQT scores. In terms of demographics, male students who enroll via distance learning are significantly less likely to successfully complete that course; however, their lower completion rate is not that different from females. Blacks have even lower completion rates in DL courses than Hispanics and whites.

Meanwhile students enrolled in courses in STEM fields are 1.9 times more likely to complete it under a traditional mode of delivery relative to DL. Those who take a Business (Humanities) course face-to-face are 2.75 (2.5) times more likely to complete it versus via DL. These results suggest that distance delivery of education is challenging regardless of the field of study.

Turning now to course grades, the mean estimates in Table 5 indicate a lack of variation in the negative effect of DL across the AFQT distribution (about 0.19 grade points lower). Women in DL do earn significantly lower grades (by 0.22 grade points lower) compared to women in resident classes, while the same gap among male sailors is slightly lower at only 0.18 grade points.

The main takeway from Table 5 is that similar to the overall QRPD estimates in Table 3, the DL-resident gap is being driven by the bottom half of the student performance distribution. This is true even when we examine QRPD estimates across different subgroups by demographics, subject fields, and AFQT terciles.

6. Conclusion

This study explicitly accounts for individual heterogeneity to provide a comprehensive analysis of the distributional effects of distance learning on student success, as measured by course grades and course completion rates. We estimate quantile effects using a rich and novel panel data of over 1.2 million courses taken by students at over 1800 institutions across the United States between 1994 and 2007. The students' setting is also unique; these courses are taken by Navy sailors as part of its Tuition Assistance program.

Given the richness of our data and the Quantile Regression for Panel Data estimator that we employ, we are able to mitigate some of the self-selection into distance courses that has plagued previous studies. For instance, students in our sample are relatively demographically homogeneous and importantly, have similar jobs and work responsibilities conditional on Rank. Our setting effectively allows us to hold constant the time constraints faced by students in comparing their outcomes. Moreover, Navy enlistees are free to choose both the courses and institutions they attend, similar to civilian college students. Their participation in DL is necessitated by largely random job rotation or deployment. Most importantly from an estimation standpoint, we exploit variation in course performance from many sailors who take multiple courses (some DL and some traditional) during the same year, in the same subject, and while serving in the same job position.

Our results confirm prior findings of negative DL effects on grades and completion rates. In addition, we find that the negative effect of DL is primarily from the bottom two-thirds of the student performance distribution. At the mean, DL delivery of education is associated with 0.19 lower course grade while at the bottom tail, the difference is up to 0.80 grade points lower. In addition, when we focus on marginal students (those who fail some of the courses that they take), we find a much larger negative effect of DL on the probability of successful course completion. Together these estimates suggest that distance learning particularly harms students who have weaker academic skills and tend to perform lower on the grade distribution. Simply estimating the mean effect of DL misses this story.

This study thus contributes to the literature on the effects of distance education on student learning by showing that the negative effects evidenced in the literature are driven primarily by students in the bottom of the student performance distribution. Our finding is also consistent with studies that found negative effects of DL are larger for students with lower prior achievement, such as those by Figlio et al (2013), Xu and Jaggars (2014), and Bettinger et al (2017). Our empirical framework extends these analyses by allowing for the heterogeneous effects of distance learning to explicitly vary across the student performance distribution rather than estimating mean effects by subgroups.

On the other hand, a major benefit of DL particularly for full-time employed students, such as those in our study, is the time flexibility it allows. Given the unpredictability in the timing of Navy sailor's job rotations, would taking a DL class, and failing to complete it or getting a lower grade be better than not taking a traditional class at all? The answer naturally depends on who it is better for.

The U.S. Navy's TA is an employer-funded education program; whether a TA student takes a college-level course via DL or face-to-face may not be as important as whether they accumulate additional human capital. A major goal of the TA program, after all, is to increase employee knowledge and skills, which should increase their job productivity.

Our estimates do show that for students at the upper end of the student performance distribution, the delivery mode of education does not matter; they perform just as well in courses that are delivered traditionally as those delivered via DL. In this sense, the promise of distance learning's increased access to higher education may be more efficiently utilized if targeted towards the more high-performing students.

Acknowledgments

We are grateful for helpful comments from three anonymous referees, the editor, and participants of the 2016 American Economic Association Annual Meetings. The views expressed in this article are our own and do not reflect the official policy or position of the Department of Defense or the U.S. government. All errors are also our own.

References

Allen, I. Elaine, and Jeff Seaman. 2016. *Online report card: Tracking online education in the U.S.* Babson Park, MA: Babson Survey Research Group.

Anstine, Jeff, and Mark Skidmore. 2005. A small sample study of traditional and online courses with sample selection adjustment. *Journal of Economic Education* 36:107–27.

Bettinger, Eric, Lindsay Fox, Susanna Loeb, and Eric Taylor. 2017. Virtual classrooms: How online college courses affect student success. *American Economic Review* 107(9):2855–75.

Bowen, William G., Matthew M. Chingos, Kelly A. Lack, and Thomas I. Nygren. 2014. Interactive learning online at public universities: Evidence from a six-campus randomized trial. *Journal of Policy Analysis and Management* 33(1):94–111.

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): 533–46.

Cosgrove, Sarah, and Neal Olitsky. 2015. Knowledge retention, student learning and blended course work: Evidence from principles of economics classes. *Southern Economic Journal* 82(2):556–79.

Defense Activity for Non Traditional Education Support, various years. Voluntary Education Fact Sheet, ed. Defense Activity for Non Traditional Education, Available http://www.dantes.doded.mil/Dantes_Web/library/docs/voledfacts.

Deming, D. J., N. Yuchtman, C. Goldin, and L. F. Katz. 2015. Can online learning bend the higher education cost curve? *American Economic Review Papers and Proceedings* 105(5):496–501.

Figlio, D., M. Rush, and L. Yin. 2013. Is it live or is it internet? Experimental estimates of the effects of online instruction on student learning. *Journal of Human Resources* 31(4):763–84.

Firpo, Sergio. 2007. Efficient semiparametric estimation of quantile treatment effects. *Econometrica* 75(1):259–76.

Jaggars, S., and T. Bailey. 2010. Effectiveness of fully online courses for college students: Response to a department of education meta-analysis. CCRC Working Paper. New York, NY: Community College Research Center, Teachers College, Columbia University.

Means, B., Y. Toyama, R. Murphy, M. Bakia, and K. Jones. 2010. *Evaluations of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Washington, DC: U.S. Department of Education.

Powell, David. 2016. Quantile regression with nonadditive fixed effects. RAND Working Paper. Available at https://works.bepress.com/david_powell/1/

U.S. Department of Education, National Center for Education Statistics. 2016. *Digest of education statistics. NCES 2016-014*. Washington, DC: National Center for Education Statistics.

Xu, D., and S. Jaggars. 2011. The effectiveness of distance education across Virginia's community colleges: Evidence from introductory college-level math and English courses. *Educational Evaluation and Policy Analysis* 33(3): 360–77.

Xu, D., and S. Jaggars. 2013. The impact of online learning on students' course outcomes: Evidence from a large community and technical college system. *Economics of Education Review* 37:46–57.

Xu, D., and S. Jaggars. 2014. Performance gaps between online and face-to-face courses: Differences across types of students and academic subject areas. *Journal of Higher Education* 85(5):633–59.