THE EDUCATIONAL IMPACT OF ONLINE LEARNING: HOW DO UNIVERSITY STUDENTS PERFORM IN SUBSEQUENT COURSES?

John M. Krieg
(corresponding author)
Department of Economics
Western Washington University
Bellingham, WA 98225-9074
John.Krieg@wwu.edu

Steven E. Henson
Department of Economics
Western Washington University
Bellingham, WA 98225-9074
Steve.Henson@wwu.edu

Abstract

Using a large student-level dataset from a medium-sized regional comprehensive university, we measure the impact of taking an online prerequisite course on follow-up course grades. To control for self-selection into online courses, we utilize student, instructor, course, and time fixed effects augmented with an instrumental variable approach. We find that students' grades in follow-up courses can be expected to be nearly one twelfth of a grade point lower if the prerequisite course was taken online. These results are robust to self-selection into online courses and into subsequent course enrollment.

doi:10.1162/EDFP_a_00196

1. INTRODUCTION

In response to the recent decline in public financial support for higher education, combined with the vast improvement in instructional technologies, universities have turned to alternative strategies to provide educational services at lower cost. The most notable of these has been to replace traditional face-to-face courses with those taught through online methods. Internet classes have become ubiquitous in higher education, with over 6.7 million students having enrolled in at least one course during the 2011–12 academic year (Allen and Seaman 2013). Thirty-two percent of higher education students now take at least one course online, and virtually all institutions with enrollment greater than 15,000 offer online courses (Figlio, Rush, and Yin 2010). There is reason to believe these numbers will increase in the future. Allen and Seaman (2013) reported 69 percent of chief academic leaders stated that online learning was critical to their long-term strategy, and 77 percent claimed online education produces the same outcomes as traditional courses.

Online instruction has become one of the most hotly debated issues in education policy today. On one side, it has been strongly supported by organizations such as the Bill and Melinda Gates Foundation and the Lumina Foundation, and it has found advocates among such influential journalists as *New York Times* columnist Thomas L. Friedman.[1] On the other side are skeptics who are concerned about the efficacy of online education and the objectives and influence of its supporters.[2] Despite the intensity of the debate about online education in the popular press, there is a surprising dearth of convincing evidence in the academic literature regarding its effectiveness. Although numerous studies have compared student performance between face-to-face and online courses, most have been descriptive studies with no controls for student self-selection into the latter. Among the most cited meta-analyses of these has been the study by Means et al. (2010). They found that although students in blended online/face-to-face classes performed modestly better than those in traditional face-to-face classes, there was no significant difference in learning outcomes between face-to-face and purely online settings. Among 1,132 papers examined, Means et al. identified only 51 studies that included control variables in a cross-sectional setting; and of these 51, only 28 studies compared online learning with face-to-face instruction. Sixteen of these 28 studies randomly assigned students to treatment and control groups, but only two (Zhang 2005; Zhang et al. 2006) had the same instructor teaching both the online and control groups. Both of these papers compared a single live lecture with a single online experience in a research laboratory—situations likely to generate different results from most settings under consideration at universities.[3]

Three more recent studies explore the impact of online instruction. Xu and Jaggars (2013, 2014) examined online course effects among students attending every community college in the state of Washington. To control for self-selection into online courses,

1. See, for example, Friedman's *New York Times* columns "Come the Revolution" (15 May 2012), "Revolution Hits the Universities" (23 January 2013), and "The Professors' Big Stage" (5 March 2013).
2. A good example is Parry, Field, and Supiano (2013). Chronologies of news articles and commentary on online education in *The New York Times* and *The Chronicle of Higher Education* can be found at http://www.nytimes.com/topic/subject/online-education-moocs and http://chronicle.com/article/What-You-Need-to-Know-About/133475/, respectively.
3. See Jaggars and Bailey (2010) for a full discussion of the U.S. Department of Education report presented in Means et al. (2010).

the authors geocoded the distance between each student's residence and their college and used this as an instrument for enrollment in online courses. They found that students enrolling in online courses were less likely to complete those courses, and they earned grade point averages (GPAs) in those courses that were about one third of a point lower than students in traditional face-to-face sections of the same courses. They further found that academic performance in online courses was worse for younger students, males, blacks, and those with lower overall GPAs.

The third study, by Figlio, Rush, and Yin (2010), randomly assigned students in a large introductory microeconomics course at a highly selective research university to either live lectures or the same lectures in an Internet setting. The authors found that course performance was higher for students in the traditional lectures, with larger positive results for Hispanic students, males, and lower-achieving students. The random assignment used by these authors dealt with self-selection into an online course, but examining a single course leaves open the possibility that these findings are driven by course-specific factors rather than the online format of the course.

All prior studies of this topic have examined the impact of online learning on course grades or some other measure of contemporaneous class performance, such as test scores or course completion. An untapped measure of performance, however, is a student's success in subsequent courses. For instance, students taking an introductory economics course online may have different outcomes when enrolling in a subsequent intermediate economics course than students who completed the prerequisite in a traditional format. Investigating the impact of online courses on follow-up course performance has a number of merits not yet exploited in the literature. For example, grades and test scores may differ between online courses and traditional courses, which potentially confounds comparisons of the two delivery methods. This problem is eliminated when comparing student performance in follow-up courses between students who took their prerequisite course online versus those who took the prerequisite in traditional face-to-face mode.

In this paper we investigate every course taught at a large comprehensive university over a ten-year period. We match each course with all subsequent courses for which it is a prerequisite. Using records from nearly 40,000 students, we compare grades earned in subsequent follow-up courses between students who completed the prerequisite in an online format versus a traditional one. Our data allow us to identify the instructor and specific prerequisite/follow-up course pair, providing the opportunity to better control for unobservables through a fixed-effects approach.

A clear pitfall of this approach is the possibility that students enrolling in online courses have unobservable characteristics that affect their success in their follow-up courses. For instance, if weaker students enroll in online courses because they perceive these courses as being easier to pass, then it would be unsurprising to observe poor performance among these students in subsequent classes.[4] Attributing this poor performance to the online prerequisite class is a clear instance of self-selection bias. Because we observe each student's outcome in other courses taken while at the university,

4. In a study of students at two Virginia community colleges, Jaggars (2014) documents a preference for students to enroll in online courses perceived as "easier" and to take courses face-to-face when the course was either "difficult" or more "important."

we are able to partly control for student-level unobservables through the inclusion of student fixed effects. As an additional safeguard against this bias, we use an instrument similar to the distance measure used by Xu and Jaggars (2013).

A second type of self-selection bias that can arise when examining follow-up course performance is that students taking online courses may have a different propensity for enrolling in follow-up courses than students taking a traditional prerequisite. For instance, students may take an online course simply to minimize their effort because the course is in an uninteresting field. Alternatively, online courses may encourage or discourage interested students from pursuing work in the academic discipline. Although we do not directly control for this type of selection, we do show that measures of a student's academic strength are uncorrelated with a student's propensity to enroll in a follow-up course after enrolling in an online course, which ameliorates concern that this type of selection influences our results substantially. A related selection issue occurs if completion of online prerequisites leads to nonrandom withdrawal from follow-up courses. We can directly test for this and determine that no such bias exists.

We find that students taking online prerequisites earn grades that are about one twelfth of a grade point lower (on a 0–4 point scale) than comparable students who took the prerequisite face-to-face. This is equivalent to about one fourth of a standard deviation in the follow-up course grade distribution, an amount slightly larger than found by Figlio, Rush, and Yin (2010). These findings are robust with respect to inclusion of student- and course-specific observables; course-, student-, instructor-, and time- fixed effects; trimming outliers from the data; and using an instrument to correct for nonrandom selection into online courses.

2. DATA

Institution Description

Data for this project come from a regional, comprehensive university that annually enrolls about 15,000 students in 160 different bachelors- and masters-granting programs, although for this project all graduate students and courses are excluded. In the most recent year, the university admitted 79 percent of applicants, about one third of whom actually enrolled. The average math and verbal SAT score of incoming freshmen is about 1,150. The university routinely is among the top five regional, comprehensive universities in the *U.S. News and World Report* rankings.

This university operates on a quarter system and typically offers many sections of the same course each quarter. This is particularly true of lower-division introductory courses, which also tend to draw larger enrollments and serve as prerequisites for upper-division courses. Most students attend three quarters per year (fall, winter, and spring). During the summer quarter, enrollment is about one fourth that of the other three quarters.

The university identifies the mode of teaching each course section as either "traditional" or "online," and both an online and traditional section of a course will often be taught during the same quarter.[5] Online courses are defined by the university as

5. Over the time period observed, the university also offered Interactive TV/video and correspondence courses, though together these accounted for 0.7 percent of all student-course observations. These were offered primarily early in the data when the university was just beginning to offer online courses. As online courses became

containing more than 75 percent of instructional time conducted via the Internet or Web-based delivery methods.[6] We define *Online* as a binary variable equal to unity if the section was taught online and zero otherwise. The university first introduced online courses in the fall quarter of 2000. Our data begin three years later (fall 2003) so the university had ample opportunity to create the infrastructure to make online courses successful by the time our observations begin. Over our sample period, 89,600 different course sections were taught, of which 1,584 (1.76 percent) were taught online. The average section enrolls 16.8 students, although the enrollment distribution is heavily right-skewed with a number of sections enrolling hundreds of students—a fact we return to in the next section.

One drawback of our data is that other than the university's classification system of online courses, we have little information describing the operation of any particular online course. Therefore, although we are sure that online courses consist of at least 75 percent Web-based material, we do not know the level of interaction between students and faculty. The university has not used multimedia specialists, content advisors, or production designers to support faculty in online courses. Instead, faculty are provided with software that records lectures and are given instruction on how to make these available to students via either an internal or external Web site. Instructors are also provided with software that allows for real-time interaction with students, though we cannot identify whether instructors utilize any form of university support. Because so much of the structure of an online course relies on the instructor, we make use of instructor fixed effects, which partly account for the variability in online course presentation between instructors.

Matched Course Structure

As with most institutions, this university delineates the course progression students need to undertake in order to enroll in upper division courses. For instance, the university requires students to take introductory microeconomics prior to both introductory macroeconomics and intermediate microeconomics. Using the official university catalog, we pair all courses with their prerequisites to create a dataset of matched courses. Using introductory microeconomics as an example, this course is paired with 27 other courses for which it serves as a prerequisite. Each matched pair consists of a prerequisite course and a subsequent follow-up course for which a student is eligible after completing the prerequisite. It is important to note that because a student is unlikely to take all follow-up courses to a particular prerequisite, many observations of the follow-up courses have missing observations for an individual student.

Using the matched course data, we identify which courses ever had a section taught online. As they do not help to identify the impact of online learning, we drop all matched courses where no section of the prerequisite course was taught online. The resulting sample contains 54 different prerequisite courses that had at least one section taught online. These 54 courses represent 3,041 different sections containing

successful, the university discontinued Interactive TV/video and correspondence courses. These were dropped from our sample to focus on traditional and online courses.

6. We verbally checked with the largest department offering online courses and learned that no content in their online offerings was presented live. We suspect most online courses present 100 percent of their material online.

Table 1. Descriptive Statistics

Panel A: Means and Standard Deviations of Online and Traditional Course Sections			
	Sections Taught Online		Sections Taught Traditionally
Course credits	3.53 (1.01)	=	3.53 (0.808)
Enrollment	21.76 (8.16)	<	50.39 (66.03)
General education	0.265 (0.435)	=	0.265 (0.443)
Number of instructors	1.01 (0.059)	=	1.06 (0.126)
GPA	3.341 (0.392)	=	3.327 (0.351)
Number of sections	351		2,690
Number of courses	54		54

Panel B: Means and Standard Deviations of Prerequisite and Follow-up Course Sections			
	Prerequisite Courses		Follow-up Courses
Proportion of sections that are online	0.115 (0.326)	>	0.061 (0.239)
Course credits	3.50 (1.13)	<	4.04 (0.884)
Enrollment per section	48.04 (63.56)	<	81.90 (65.83)
General education	0.251 (0.435)	<	0.516 (0.498)
Number of instructors	1.05 (0.122)	=	1.05 (0.122)
GPA	3.327 (0.339)	>	3.066 (0.370)
Number of sections	3,041		3,317
Number of courses	54		187

Note: >, < indicate statistical significance at the 95% level.

1,080,325 student–section observations. Of these 3,041 sections, 351 (11.5 percent) were taught online.[7] The 54 prerequisite courses were matched with 187 different follow-up courses representing 3,317 different sections and 66,164 student-section observations. Of these 187 follow-up courses, all but seven were taught in the same department as the prerequisite, suggesting the linkage between the characteristics of the prerequisite course and performance in the follow-up should be more closely related to performance in the prerequisite than to such factors as a student's motivation or interest in the subject matter. Because many students may take multiple prerequisite courses that appear in the matched data, we observe each individual student in an average of 19.9 follow-up courses, which we later exploit in a student fixed-effects framework.

For each of the 54 prerequisite courses having both online and traditional sections, panel A of table 1 presents averages across sections within each instructional mode.

7. As stated earlier, this university has offered 1,584 different online sections. Note, however, that 1,233 of these lacked a follow-up course and were dropped from our sample.

Other than their method of instruction, online and traditional courses fill the same university requirements, which is why there is no difference in the number of credits offered or the percentage of sections that fulfill general education requirements. The only statistical difference between online and traditional sections is that traditional sections are on average more than twice the size of those taught online. Traditional sections average an enrollment of 50.39 students, whereas online sections average 21.76 students. This can be partially explained in that our observation period includes the early years of online education when students were more reluctant to enroll in online courses. Interestingly, there is no statistically significant difference in average section GPA by course type; both online and traditional prerequisites average GPAs slightly better than a B+ at 3.34 and 3.32, respectively.[8]

Panel B of table 1 displays descriptive statistics for prerequisite and follow-up courses in our sample. The 11.5 percent of sections in prerequisite courses that are taught online is nearly twice the online proportion of follow-up course sections, a fact that is explained by the university's use of online courses in introductory courses. Follow-up courses are larger (by enrollment), are more likely to be taught as part of the university's general education program, and are offered for more credits than are prerequisites. Prerequisite courses average higher GPAs than do follow-ups.

Student Data

After dropping students who do not appear in any matched course with an online prerequisite and those with missing data, our sample consists of 38,652 undergraduates attending the university between fall 2003 and spring 2013. The average student enrolled for 9.0 quarters, yielding 348,881 student–quarter observations.[9] Importantly, we are able to observe student information that changes over quarters, including credits enrolled during each quarter, cumulative credits earned, age, and local housing location. In addition, for each student we observe the set of sections taken and the grade earned in each. The section data also contain a unique variable identifying the section's instructor. The section data include 1,230,993 student-section records, for an average of 31.8 sections taken per student.

As noted earlier, one concern is that students enrolling in online courses might differ in unobserved ways from those who do not. If students nonrandomly select into online courses, then simple comparisons of outcomes will generate biased results. We explore this possibility in table 2, which separates students into two groups: those who took at least one online course, and those who took none. In contrast with the findings of Xu and Jaggars (2013), who studied two-year college students, students at this university who took at least one online course are on average weaker, as measured by both SAT math score and high school GPA (although these differences are not large). On the other hand, students who took an online course averaged a higher first-quarter GPA at this institution than those who did not. Part of this may be explained by the fact

8. GPAs at this university are on a four-point scale with an option for instructors to assign pluses and minuses for all grades except As, which can only receive minuses. For example, a B+ is recorded as a 3.3 GPA and a B− is recorded as a 2.7.

9. This is below the number of quarters it typically takes to graduate because we observe students who transferred into the university and hence take less time to graduate. We also included observations censored at both the beginning and ending of our observation time.

Table 2. Means and Standard Deviations of Students

	Students Who Took at Least One Online Course		Students Never Taking an Online Course
First generation college student	0.344 (0.475)	=	0.343 (0.473)
Ethnic minority	0.191 (0.393)	=	0.188 (0.391)
Verbal SAT	552.8 (83.14)	=	556.4 (84.51)
Math SAT	548.1 (78.51)	$<$	554.3 (77.91)
High school GPA	3.43 (0.434)	$<$	3.47 (0.348)
Age upon entry, years	22.97 (8.18)	$<$	25.10 (10.17)
Credits upon entry	35.9 (42.9)	$>$	26.2 (37.0)
First quarter GPA	3.05 (0.716)	$<$	2.96 (0.699)
Number of students	2,875		35,777

Note: $>$, $<$ indicate statistical significance at the 95% level.

that students enrolling in an online course average a higher number of transfer credits brought to the university and therefore may take a different mix of courses during their first quarter. Students never taking an online course are also older, perhaps reducing the signal value of SAT math scores and high school GPAs.

A similarity with Xu and Jaggars (2013), unreported in table 2, is that students enrolling in online courses are less likely to complete these courses than are students in traditional courses. This university counts the number of course withdrawals that occur after the sixth week of the ten-week quarter. Among students registered in an online course, 6.9 percent drop the course after the sixth week, whereas 3.0 percent of students in traditional courses do so.[10] Although we have no method of measuring drops prior to the sixth week, if weaker students are more likely to drop courses, then those who complete online courses would be relatively stronger students than those who complete traditional courses, thus leading to an upward bias in estimates of the impact of online courses on future performance.

Ultimately, we are interested in the impact that an online prerequisite course has on students' grades in subsequent follow-up courses. A cursory method of answering this question is to compute the average grade earned in follow-up courses by prerequisite course type. The average follow-up grade for students who took a traditional prerequisite is 3.135, whereas the average follow-up grade for students who took an online prerequisite is 3.455, a difference that is large both practically and statistically ($t = 18.55$). These average differences are confounded by a number of factors, however, including omission of explanatory variables, possible nonrandom selection of students into online courses, nonrandom selection of follow-up courses, and the possibility that

10. This university computes GPA as the number of grade points earned divided by the number of credits attempted. A course that is dropped adds zero grade points earned and zero credits attempted, resulting in no impact on a student's GPA.

online courses are offered in departments or by instructors that give higher than average grades.[11] The next section describes how we overcome these possible difficulties.

3. METHODOLOGY

Using the matched course data, we compare the performance in follow-up courses of students who took the prerequisite online versus those who took it traditionally. Consider student i who takes course j in a series of courses for which course $j = 1$ is a prerequisite. Course $j > 1$ is taught by instructor k in quarter t. We model this student's grade G in course $j > 1$ as:

$$G_{ijkt} = \beta\, Online_{i1} + \mathbf{\Gamma X}_{ij} + \mathbf{\Delta Z}_{i1} + \mathbf{\Lambda W}_{ijt} + u_i + v_k + t_t + e_{j1} + \varepsilon_{ijkt}. \tag{1}$$

The parameter β identifies the impact of taking the prerequisite course $j = 1$ on the grade earned in the follow-up course j.[12] In this case G is measured as the number of grade points earned on a 0 to 4 scale. Equation 1 contains two types of error terms (the ε_{ijkt}), which represent the unobserved grade component and four fixed-effect terms. These are u_i, which represents student-specific unobserved variables; v_k, which is an instructor-specific error term; t_t, which indicates the year-quarter in which the course was taught; and e_{j1}, which represents unobserved effects associated with a particular prerequisite/follow-up course pair. The e_{j1} account for the possibility that prerequisite $j = 1$ might have a different effect on one follow-up course than it does on another. Thus, each matched pair has a unique e_{j1}. We estimate equation 1 as a four-way fixed-effects model. Because of the matched-course structure of the data, in which a particular student may have multiple follow-up courses corresponding to a given prerequisite course—thus violating the assumption of independent observations—standard errors are corrected for clustering at the prerequisite course level of each individual student.

As students enter the follow-up course having come from either a traditional or an online prerequisite, we could potentially identify β as the difference in follow-up grade between students with these two backgrounds. Doing so, however, would ignore both the inherent differences between online and traditional prerequisites described in table 1, and the differences among students taking online courses described in table 2. To account for these effects, we include in equation 1 a number of control variables that help explain follow-up course grade. The $\mathbf{X}_{ij}$ variables are specific to student i's experience in section j. They include indicators for whether follow-up section j was taught online, whether it satisfies a general education requirement, and whether it was taught in a traditional lecture format or identified by the instructor as a seminar course. The $\mathbf{X}_{ij}$ also include section j's enrollment. The $\mathbf{Z}_{i1}$ represent characteristics of the prerequisite course, including the grade earned by the student in that course, its enrollment, and whether it satisfies a general education requirement. The $\mathbf{W}_{ijt}$ represent student characteristics associated with taking course j at time t. These include the time (measured in quarters) and its square since completing the prerequisite

11. The three academic departments that have offered the most online courses are Elementary Education, Secondary Education, and Human Services, which are among the highest grading departments on campus.
12. Although an ordered logit model may be more appropriate in accounting for the discrete nature of course grades, ordered logit models are well known for their inability to accommodate a fixed effects structure; thus we estimate equation 1 by least squares.

Table 3. Fixed Effects Estimates of Course Grades, Selected Coefficients

Variable	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) 2SLS
Online_{i1t}	−0.083**	−0.071**	−0.073**	−0.074**	−0.314**	−0.079**	−0.071**	−0.092
	(0.035)	(0.034)	(0.035)	(0.035)	(0.149)	(0.037)	(0.035)	(0.070)
Online_{ijt}	−0.086	−0.057	−0.058	−0.055	−0.054	−0.089**	−0.055	−0.047
	(0.053)	(0.053)	(0.053)	(0.053)	(0.053)	(0.041)	(0.053)	(0.052)
Both Online_{ij1t}	−0.022	0.010	0.014	0.014	0.007	0.032	0.004	−0.046
	(0.068)	(0.067)	(0.067)	(0.067)	(0.067)	(0.072)	(0.067)	(0.059)
Grade_{i1t}	0.024***	0.024***	0.022***	0.023***	0.020***	0.074***	0.024***	0.022***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.010)	(0.006)	(0.006)
AvgGPA_{ijt}		0.436***	0.437***	0.437***	0.436***	0.587***	0.437***	0.437***
		(0.016)	(0.016)	(0.016)	(0.016)	(0.022)	(0.016)	(0.016)
$\text{OnlineExperience}_{i1t}$			−0.0003	−0.0003	−0.0003	−0.0005	−0.0004	−0.046
			(0.0008)	(0.0008)	(0.0008)	(0.0009)	(0.0008)	(0.058)
$\text{DeptExperience}_{ijt}$				−0.003*	−0.003*	−0.0004		−0.003*
				(0.002)	(0.001)	(0.0004)		(0.002)
$\text{Online}_{i1t} \times \text{Grade}_{i1t}$					0.072*			
					(0.043)			
Hausman test of endogeneity								$t = -1.6$
No. of Student-matched Courses	66,164	66,164	66,164	66,164	66,164	21,677	48,400	66,164
No. of Students	16,738	16,738	16,738	16,738	16,738	11,225	12,264	16,738
No. of Follow-up Instructors	300	300	300	300	300	148	250	300

Notes: Standard errors corrected for clustering at the matched course/student-level presented in parenthesis. All models include student fixed effects, matched course fixed effects, instructor fixed effects, and quarter fixed effects. 2SLS = two-stage least squares.

*Statistically significant at the 90% level of confidence; **statistically significant at the 95% level of confidence; ***statistically significant at the 99% level of confidence.

course; the student's age; a polynomial in the number of accumulated credits earned at the time of enrolling in course j; and binary variables identifying whether course j and its prerequisite $j = 1$ were taught by the same instructor, whether both course j and its prerequisite were taken online, and whether course j is in the student's major area of study.

4. RESULTS

Online Impacts on Follow-up Course Grade

Estimates of β and other selected coefficients from equation 1 are presented in the first column of table 3. After taking an online prerequisite, the performance of students in a follow-up course is expected to decline by 0.083 grade points—an estimate that is both statistically and practically significant.[13] The magnitude of this estimate suggests that a student taking an online prerequisite course is expected to have their follow-up course grades decline by roughly one twelfth of a full letter grade, or about one fourth of the difference between fractional grades (e.g., an A versus an A–). This estimate is roughly three times as large as the conditional impact of a one-point change in

13. In the discussion of table 2 we note average follow-up grades of 3.135 and 3.455 for students who took a prerequisite in traditional and online modes, respectively—a statistically significant difference of 0.32 points in favor of online prerequisites. This raw difference ignores the effects of observed and unobserved differences between the two groups. Adding the full set of explanatory variables reduces this difference to zero. Adding the fixed effects, either individually or in any combination, results in a negative impact of online prerequisites. The numerical value of the estimated β is quite robust with respect to the specification of the fixed effects. A sensitivity analysis addressing this point is available from the authors on request.

the student's grade in the prerequisite course and about 60 percent larger than the (unreported) impact of having the same instructor teach both the prerequisite and follow-up sections. Another way of putting this estimate into context is by comparing it to the (unreported) 300 instructor fixed-effects estimates generated by equation 1. The estimated β of –0.083 corresponds to the 34th percentile of the distribution of instructor fixed effects, suggesting that taking an online prerequisite is roughly equivalent to enrolling in a class taught by an instructor who assigns grades at the lowest one third of the university distribution versus the median. A final method of understanding this effect is in terms of grade standard deviations. As reported in panel B of table 1, the average follow-up GPA is 3.066 with a standard deviation of 0.37. A reduction of 0.083 grade points is equivalent to 0.22 standard deviations, an amount slightly larger than the 0.17 standard deviation[14] Figlio, Rush, and Yin (2010) found when they examined the impact of online instruction on contemporaneous learning.

Robustness Checks

There are a number of threats to identifying the causal impact of taking an online course on follow-up course performance. In this section, we investigate omitted variable bias and the adequacy of the student-level fixed effects structure. We demonstrate that the student-level fixed effects sufficiently control for student-level unobservables that might confound estimates of β. We also present alternate formulations of equation 1 that include specific variables whose exclusion could bias estimates of β.

It is possible that students who take online prerequisites sort into follow-up sections that assign easier grades. Although this is somewhat controlled for by the inclusion of instructor fixed effects, a stronger control would be the average grade earned by students other than student i in section j ($AvgGPA$).[15] If students sort into courses based upon expected ease of grading, then the inclusion of $AvgGPA$ should capture much of this. Column 2 of table 3 adds this variable to the regressors of equation 1 with the result that $\hat{\beta}$ falls in magnitude to –0.071, about one thirteenth of a grade point, but remains statistically significant.

A second possible source of bias relates to the instructor's experience teaching online courses. If an instructor is relatively new to teaching online, then students' poorer performance in subsequent courses might result from the prerequisite instructor's inexperience with the format rather than from the format itself. This is partly mitigated by the fact that online courses at this university have been taught since 2000. Thus there is ample institutional experience with online courses, though not necessarily individual faculty experience, by the time this sample begins. To control for this possibility we create the variable *OnlineExperience* which, for each observed quarter in the data, counts the cumulative number of sections previously taught online by instructor k.[16] We interact *OnlineExperience* with *Online* to allow an instructor's online teaching experience to have a differential impact on follow-up course grade depending on whether the prerequisite course was taught online or in traditional face-to-face mode. We add

14. Computed by Krieg and Henson.
15. Specifically, for section j we compute the average grade earned by all students other than student i and assign that average to student i.
16. For instructors with experience teaching online courses prior to the beginning of our observations, this measure will understate their experience teaching online courses.

these variables to equation 1 and report the results in column 3 of table 3. The inclusion of these variables does little to the estimate of β, which remains about one thirteenth of an overall grade point and statistically significant.[17]

Another possibility is that students entering a follow-up course have different amounts of experience within an academic discipline. If these differences are correlated with their online background, then the prior regressions may confuse online prerequisites with familiarity and experience in a discipline. For instance, if students taking online courses are less attached to a discipline, then they may have taken fewer related courses by the time they enroll in a follow-up course. Their lack of experience may then reduce their follow-up course grade. To control for this, we define *DeptExperience* as the cumulative number of courses in an academic department completed by student i prior to enrolling in a follow-up course offered by that department. The inclusion of *DeptExperience* does little to change the estimate of β which, as shown in column 4 of table 3, increases in magnitude slightly to –0.074.

Another concern has to do with the possibility that grades earned in online courses may have a different informational content than do those earned in traditional courses. All of the preceding models have used the grade earned in the prerequisite ($Grade_{i1t}$) as an explanatory variable. If online courses assign grades differently than do traditional courses, then $Grade_{i1t}$ will have a differential impact on follow-up course grades based upon the type of prerequisite completed. For instance, if online courses assign grades more leniently, then an online student would be less well prepared for a follow-up course than a student who completed the traditional prerequisite and earned the same grade, generating a downward bias in $\hat{\beta}$. To account for this possibility, we interact $Grade_{i1t}$ with $Online_{i1t}$ and include it in equation 1. In essence, this interaction allows the prerequisite grade to have a different impact on the follow-up course depending upon whether the prerequisite course was taken online or in traditional mode.

Results including the interaction of these two variables appear in column 5 of table 3. The *Online* coefficient drops dramatically to –0.314 but the interaction term is positive, suggesting a differential impact of online courses depending on a student's performance in the prerequisite.[18] Among students who earned high grades in prerequisite courses, the difference in follow-up GPA between those who took online and face-to-face prerequisites is minimal. For example, a student earning an A (or A–) in an online prerequisite can expect to earn a follow-up GPA that is 0.026 (or 0.046) grade points lower than a traditional student, respectively, though neither of these estimates differs statistically from zero (*p*-values of 0.560 and 0.212). In contrast, students who received lower grades in the prerequisite course earned significantly lower follow-up course grades if their prerequisite was completed online. For instance, a student earning a prerequisite online course grade of B could expect his follow-up grades to drop by 0.098 grade points, a statistically significant amount ($p = 0.018$). A student earning a C in the same online prerequisite could expect a 0.170 drop in grade points ($p = 0.017$)

17. Not listed in table 3 is the coefficient on the product of *OnlineExperience* and *Online*. This coefficient is –0.0006 ($p = 0.712$).

18. Ignoring other terms, the predicted grade in the follow-up course is $G_{ijkt} = -0.314(Online_{i1t}) + 0.020(Grade_{i1t}) + 0.072(Online_{i1t} \times Grade_{i1t})$, so that for a given value of $Grade_{i1t}$ the predicted grade difference in the follow-up course between those who took the prerequisite online versus in traditional mode is $\Delta G_{ijkt} = -0.314 + 0.072(Online_{i1t})$.

relative to a face-to-face prerequisite. At least two explanations for this grade differential are possible. First, as noted above, online courses may assign grades more leniently, resulting in lower follow-up grades for students earning similar prerequisite grades. Second, it is possible that online prerequisites have differential impacts based upon student ability.[19] Since our only measure of students' past performance in a prerequisite is their grade in that course, the data do not allow us to distinguish between these two possibilities.

Another concern with equation 1 is that the inclusion of student-level fixed effects may inadequately control for unobservables related to course performance. To check this, we perform the following exercise. For each prerequisite taken by a student, we randomly select another course completed by that student and use it in place of the follow-up course previously used. We constrain the random selection of these courses to those taken during a different quarter than the prerequisite (to prevent temporal correlation between grades) and to those taught outside the prerequisite's academic department (to prevent grade correlation caused by similarity of subject). We then estimate equation 1 using these new, randomly matched data. In essence, this test amounts to estimating the impact that taking an online course in one subject has on the grade earned in a completely different subject during a different quarter. If our control variables (including student-level fixed effects) adequately control for student academic ability, then we would expect taking an online course has no impact on unrelated follow-up courses. In other words, we would expect β from this experiment to equal zero. Deviations from $\beta = 0$ would imply that the matched course structure may be generating erroneous results because we have inadequately controlled for academic background or innate student ability. Because there are many possible randomly matched courses for each student, we replicate this experiment 500 different times and graphically present the estimated βs in figure 1. The average $\hat{\beta}$ from these simulations is 0.0047 and is not statistically different than zero ($p = 0.128$). Nor are there frequent instances of these randomly assigned course pairs generating estimates of β near those produced in table 3: Only eight of the 500 estimates produced a $\hat{\beta}$ more negative than –0.071, the closest ordinary least squares (OLS) estimate to zero in table 3, and only one of them was further from zero than –0.083, the most negative value in that table. Taken as a whole, this experiment suggests that our online estimates are not driven by bias in the matched course structure of the data or by a lack of adequate control variables.

On the other hand, it is possible that these results may understate the true impact of online courses on follow-up grade. The matched structure of the data is such that one prerequisite can lead to many follow-up courses, even though some of those follow-up courses build upon intervening ones. For instance, to take course #3, a student must complete courses #1 and #2, but #1 is a prerequisite to #2. The matched course-data approach links #1 to #3 yet it is reasonable to imagine that by the time a student takes

19. This would be consistent with Figlio, Rush, and Yin (2010), who found that the performance gap between weaker students (defined as students with collegiate GPAs below the median) and stronger students tended to be larger in online sections of a given course. Xu and Jaggars (2014) made a similar finding when studying community college students in the state of Washington. In their analysis, online education resulted in lower academic performance among men, younger students, black students, and students with lower grade point averages.

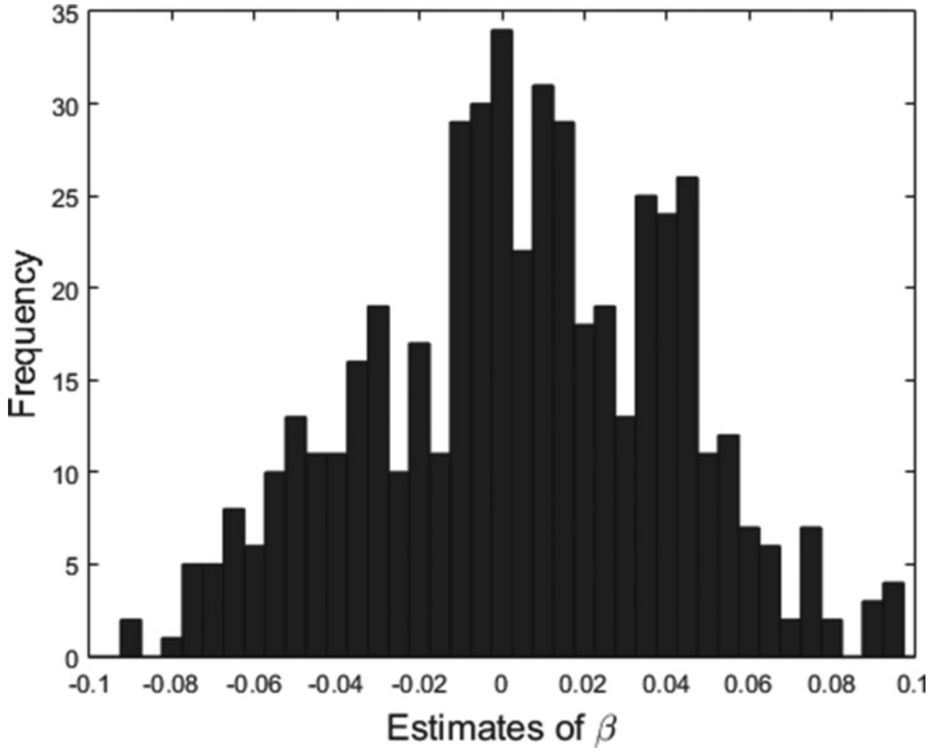

Figure 1. Distribution of $\hat{\beta}$ from Randomly Matched Course Pairs.

an advanced class like #3, the impact of #1 on #3 has been diluted by the intervening course #2. If that is the case, then one would expect to find the average impact of an online prerequisite to be smaller than its impact on the most immediate follow-up course.

To overcome this, we revisit the matched data and select a single follow-up course for each prerequisite rather than considering all possible follow-up courses. For this exercise, we limit the analysis to the follow-up course with the closest subsequent course number to the prerequisite in the same department.[20] Limiting the data in this way reduces the number of observations substantially, but it also more closely links prerequisites to follow-up courses in the way that students would naturally enroll in them, and focuses on the impact of an online prerequisite on subsequent performance in the same discipline.

Using the subsample of matched-course data, we re-estimate equation 1 and report the results in column 6 of table 3. As expected, in the subsample the magnitude of $\hat{\beta}$ increases by about 10 percent to –0.079, suggesting the negative impact of online courses is slightly larger for courses “closer to” the prerequisite in terms of both sequencing and subject matter.

20. As an example, Psychology 102, and not Psychology 105, would be the closest follow-up course to Psychology 101.

Selection Issues

Two types of selection issues can confound a causal interpretation of β. First, students may nonrandomly select into online prerequisites. For instance, if weaker students enroll in online courses, then finding their grades are lower in follow-up courses would not be sufficient to conclude that online courses reduce follow-up performance. Second, relative to completing a face-to-face prerequisite, after taking an online prerequisite students may differentially select into follow-up courses. The two following subsections deal with the first type of selection and the third subsection examines the impact of nonrandom selection into follow-up courses.

Selection into Online Courses Based upon Observables

Table 2 documents that students who take an online course are different from those who do not, as measured by high school GPA, math SAT scores, first-quarter collegiate GPAs, age, and credits upon entry. Although the approach of equation 1 implicitly controls for this by using student-level fixed effects, one concern is that our estimates are a product of differences in academic background rather than the impact of online courses. As a check for this, we follow the trimming procedure introduced by Crump et al. (2009) and described by Imbens (2014) (we refer to this as the CHIM procedure). This procedure involves estimating a propensity score for each student and then trimming the sample to make the range of propensity scores of nontreated individuals similar to the range of treated individuals. This differs from a true matching model in that rather than matching individuals based on their propensity scores and then estimating average differences in outcomes, the CHIM procedure simply drops individuals with propensity scores far from the common support and then re-estimates the econometric model with the trimmed sample. This approach is motivated by the idea that if there is a value for the covariates such that there are few treated units relative to the control units, then these covariates will have a disproportionately larger impact on the estimates of β. Excluding these units should improve the asymptotic variance and limit the impact of implicit functional form assumptions on the estimator. Another benefit of the CHIM procedure relative to a more traditional matching model is that it can be applied to individuals based on the unchanging characteristics listed in table 2, in contrast with a course-by-course approach that would be required for matching students within prerequisite/follow-up course pairs.

To operationalize the CHIM procedure, we utilize a logit to estimate equation 2:

$$\Pr(Enroll_i = 1) = \mathbf{\Psi H}_i + \mathrm{w}_i, \tag{2}$$

where *Enroll* is a binary variable equal to one if student *i* ever enrolled in an online course and zero otherwise. The matrix **H** was determined by the Imbens and Rubin (2015) procedure for choosing independent variables in the propensity score estimation.[21] In this case, **H** contains all of the time-invariant student variables presented in table 2 as well as the squares of the number of credits transferred into the university and the student's first-quarter collegiate GPA.

21. The Imbens/Ruben procedure is a stepwise specification search that systematically includes polynomials and interaction terms of the independent variables.

Using the predicted estimates from equation 2, we calculate each student's propensity of ever enrolling in an online course. We then follow the CHIM trimming procedure, which identifies extreme values of the propensity score that are dropped from the analysis. The trimming eliminates 4,474 individual students representing 26.7 percent of the total student sample. After trimming the sample, we re-estimate equation 1 and reproduce the results in column 7 of table 3. The estimated impact of enrolling in an online prerequisite is stable at –0.071 with standard error similar to those in the regressions based upon the full sample. The similarity in estimates of β between the trimmed and full samples suggest the online findings are not driven by systematic differences in student academic ability measured by their high school performance, their first quarter on campus performance, math SAT scores, or measurable demographic differences.

Student Selection into Online Courses Based on Unobservables

A significant drawback of earlier studies of online courses has been the failure to account for student-level unobservable influences that may simultaneously affect both a student's willingness to enroll in an online course and their performance. For instance, if stronger students are more likely to enroll in online courses, as Xu and Jaggars (2013) suggest, then it would not be surprising for researchers to erroneously estimate positive impacts of online experiences on subsequent coursework. At this university it appears that marginally weaker students (as measured by math SAT and high school GPAs presented in table 2) are more likely to enroll in online courses. Hence, failing to account for unobserved student ability may cause the opposite error of falsely concluding that online courses decrease subsequent performance, and this could potentially explain the negative estimates of β in the first seven columns of table 3.

A number of features of equation 1 mitigate against this possibility. Equation 1 includes student-level fixed effects, which account for time-invariant unobserved student-level factors, such as innate academic ability. In light of this, β is interpreted as the impact of taking online course 1 on the grade in course j, holding constant unobservable student-level factors that influence the grade. To the extent these unobservable student-level factors are correlated with the likelihood of enrolling in an online prerequisite, we already control for this self-selection. A related concern is that students who are more capable enroll in subjects that are more difficult and that may assign grades in a manner different than in easier subjects. The inclusion of the matched-course fixed effect term inherently accounts for these types of differences that occur in course-pairs.[22] Likewise, students may nonrandomly choose instructors—this, too, is partially accounted for by including instructor fixed effects.

Although individual fixed effects mitigate the impacts of student self-selection into different courses, there may be a remaining selection issue if students sort between online and traditional courses on a course-by-course or term-by-term basis. To deal with this concern, we follow Xu and Jaggars (2013) by introducing an instrument for $Online_{i1}$. Specifically, for each follow-up course, we compute the distance to campus from the student's local residence during the quarter in which the student enrolled

22. In addition, dummy variables indicating the student's major area were added to help control for differences in course grading across departments. The addition of these variables did not meaningfully alter the estimates of *Online*.

in the prerequisite course ($Distance_{i1t}$). *Distance* is computed by using GIS systems to measure the distance in meters between the campus' central administration building and each student's local mailing address.[23] Specifically, we estimate:

$$Online_{i1t} = \sum_{q=1}^{3} \delta_q \, Distance_{i1t}^{q} + \mathbf{\Gamma X}_{ij} + \mathbf{\Delta Z}_{i1} + \mathbf{\Lambda W}_{ijt} + u_i + v_{jk} + t_t + e_{j1} + v_{ijkt}. \tag{3}$$

In equation 3, we use a cubic in distance because of the possibility that online course selection is nonlinear in distance and because a cubic minimizes the Akaike information criterion of equation 3. Because of the well-known drawbacks[24] of including fixed effects in probit and logit models, we estimate equation 3 with a linear probability model.

For *Distance* to be a valid instrument it must explain *Online* in equation 3 and simultaneously be uncorrelated with ε_{ijkt} in equation 1. As to the first condition, the three coefficients on the cubic of *Distance* from equation 3 (with *t*-statistics in parentheses) are: 5.2e–8 (5.44), –1.7e–14 (–2.20), and 2.16e–21 (1.67). The first two coefficients are statistically significant at any conventional level of significance, and the third is significant at the 90 percent level. Jointly, these three are strongly significant (*F*-statistic of 29.60, p-value = 0.0000).

As to the second condition, there are at least two potential threats to the validity of using distance as an instrument: (1) those who value education may choose to live closer to campus, and (2) students living closer to campus might perform better because of lower travel cost and access to campus resources.[25] Either scenario would lead to a correlation between *Distance* and the ε_{ijkt} in equation 1.[26] Despite these concerns, there is little evidence that distance from campus affects GPAs directly. Schrager (1986) found no evidence of such an effect; and in a recent study of the impact of living on campus freshman year, the University of California, Irvine, found no relationship between GPA and whether a student lived on or off campus (UCI 2007). Some researchers (e.g., Card 1995; Rouse 1995) have suggested that families who value education choose to live near a college campus, though it is important to point out these researchers are concerned with the degree of proximity chosen by a student's family rather than with the choice made by the student after enrolling in a particular campus.

To more formally address these concerns, we follow Xu and Jaggars (2013) by conducting a falsification test to assess the relationship between course outcomes and distance for traditional courses. Specifically, we exclude all online courses from our sample and, using this subsample, we re-estimate equation 1 including the cubic in distance but without the online observations and associated variables. Omitting the online observations isolates the potential effect of *Distance* on grade and prevents

23. The mailing address differs from the student's permanent address. The mailing address is where the university sends student mailings over the course of the quarter, whereas the permanent address is usually the location of the student's parents.
24. See Greene (2001) or Heckman (1981).
25. We realize that there may be other arguments leading to *Distance* being correlated with the error term, some of which may even entail better students living farther from campus. Our intent is to demonstrate that this correlation is potentially an issue and to elucidate how we measure it.
26. This problem is somewhat mitigated in that *Distance* is measured during the quarter the student enrolls in the prerequisite course and the ε_{ijkt} occur during the quarter the student takes the follow-up course.

Distance from impacting grade through selection of online courses. If students living farther from campus were systematically less motivated or receive fewer institutional resources, then distance would be inversely related to grades in this sample. The results of this experiment suggest distance has no bearing on course grade. None of the coefficients in the cubic of *Distance* was individually significant, nor were they jointly significant ($F = 1.26$, $p = 0.261$).

Given these results, we re-estimate equation 1 using two-stage least squares with a cubic in *Distance* as the instrument. Column 8 of table 3 replicates those of the fourth column of the same table with the exception that *Distance* instruments for $Online_{i1t}$. In this case the magnitude of the estimated β has grown, suggesting that OLS may underestimate the negative impact of online courses. Nonetheless, as one might expect when using an instrument, the standard errors associated with $\hat{\beta}$ double, leaving the *Online* estimate statistically insignificant. Given the instrumental variable estimator is consistent but less efficient than OLS when there is no correlation between the independent variables and the error term, one possibility is that in the presence of individual-level fixed effects, there is no correlation between *Online* and the error term of equation 1. As an example, if weaker students are more likely to enroll in an online course and earn lower grades in follow-up courses, the inclusion of student-level fixed effects controls for academic ability, leaving *Online* and the ε_{ijkt} uncorrelated. We explore this with a Hausman test of endogeneity for the instrumented column in table 3 and find no evidence of correlation between the error term and *Online*. We conclude that instrumenting for *Online* is unnecessary and OLS provides unbiased and efficient estimates.

Selection into Follow-up Courses

Unlike prior research that examines the effect of online courses on outcomes in that course, the use of the matched course structure introduces a second possible route for nonrandom selection in that students taking online prerequisites may be more or less likely to choose follow-up courses than those taking a traditionally taught prerequisite. One piece of evidence supporting this possibility is that students completing a traditional prerequisite enroll in 20 percent more follow-up courses than do those completing online prerequisites. If taking an online prerequisite discourages academically strong students from pursuing additional courses in that discipline, then the GPA results of table 3 would compare relatively weaker students coming from an online prerequisite with stronger traditional students, leading to a negative bias in $\hat{\beta}$. Of course the opposite effect could occur, in which online courses discourage weak students from taking related follow-up courses, leading to an upward bias in estimates of β. Traditional approaches to correcting for this type of selection require an instrument that is correlated with the choice to pursue follow-up courses but uncorrelated with the unexplained component of grades. Unfortunately, our data do not contain a suitable instrument. Instead, we observe measures of general academic ability and demonstrate that these are uncorrelated with enrollment in a follow-up course after a student takes an online prerequisite.

To investigate what type of students enroll in follow-up courses, consider the following equation:

$$\text{Prob}[Enroll_{ij1t} = 1] = \boldsymbol{\alpha}\mathbf{X}_i + \boldsymbol{\lambda}\mathbf{Z}_{i1} + \boldsymbol{\Lambda}\mathbf{W}_{i1t} + e_{j1} + \varepsilon_{ijkt}, \tag{4}$$

Table 4. Linear Probability Estimates of the Impact of Student Characteristics on Enrollment in Follow-up Courses

Variable	(1)	(2)	(3)
High school GPA	−0.012 (0.007)		−0.006 (0.008)
SAT Verbal		−0.00001 (0.00003)	−0.000006 (0.00003)
SAT Math		−0.00003 (0.00004)	−0.00004 (0.00004)
Matched course fixed effects	YES	YES	YES
No. of student-matched courses	26,659	26,659	26,659
No. of students	2,875	2,875	2,875

Note: Standard errors weighted for linear probability heteroskedasticity and corrected for clustering at the student-level presented in parentheses.

where $Enroll_{ij1t}$ is a binary variable equal to one if student i enrolls in prerequisite course 1 and its follow-up course j. In order to investigate the types of students who enroll in follow-up courses after taking an online prerequisite, when estimating equation 4 we limit the sample to only students who enrolled in an online prerequisite. As in equation 1, the matrix **Z** contains variables that describe the prerequisite course[27] and matrix **W** contains information about student i when they enrolled in the prerequisite course.[28] The error term e_{j1} represents specific fixed effects for each prerequisite/follow-up course pairing.

In equation 4 the variables of interest are included in **X**, which contains time-invariant student characteristics. The time-invariant student characteristics we observe are limited to the student's high school GPA, as well as their math and verbal SAT scores.[29] We estimate three variations of equation 4 and present these results in table 4. In order to exploit the panel nature of the data, we utilize a linear probability model with fixed effects for each prerequisite/follow-up course pair and weight the standard errors to account for the heteroskedasticity that arises in linear estimates of binary choice models.

The three columns of table 4 represent model specifications that contain different combinations of high school GPA and SAT scores. In all specifications, high school GPA and SAT scores fail to predict whether a student is more or less likely to enroll in a follow-up after taking an online course. This suggests that there is little reason to be concerned about nonrandom selection of students into follow-up courses based upon their academic ability.

Although table 4 demonstrates that the choice to enroll in follow-up courses is orthogonal to high school GPA and SAT scores, it remains possible that the GPA results of table 3 are influenced by nonrandom completion of follow-up courses. Consider the possibility that students withdraw from follow-up courses in a pattern correlated

27. These include the prerequisite's enrollment, its average GPA given to students other than student i, whether the course was a general education requirement, and the term in which it was offered.
28. These include whether the course was offered in student i's major area of study, the grade earned in the prerequisite, the student's age, a quartic in the student's accumulated credits up to that time, the grade earned in the prerequisite course, and a binary variable indicating whether the prerequisite was in the student's major.
29. Also included in **X** are six binary variables representing the student's ethnicity, gender, age upon enrollment, and whether the student was a resident of the same state that contains the university.

with the delivery method of their prerequisite. With respect to the results of table 3, two particularly concerning cases exist: stronger students who completed a traditional prerequisite course or weaker students who completed an online course. In either case, if these students were more likely to withdraw from follow-up courses, one would expect to bias $\hat{\beta}$ downward and generate the finding that online courses are less effective than traditional ones.

We begin our investigation of nonrandom withdrawal from follow-up courses by noting that, conditional upon registering for a follow-up course, 1.95 percent of students who enrolled in an online prerequisite end up withdrawing from the follow-up. For those completing a traditional prerequisite, 2.49 percent withdraw from the follow-up, a small difference that is not statistically significant ($p = 0.116$). We formalize this by estimating a variant of the four-way fixed effects equation 1 in which we utilize the dependent variable *Withdraw*, which is equal to one if a student withdraws from a follow-up in a matched course pair. This approach amounts to a linear probability model of *Withdraw* with the advantages of accounting for time, matched course pair, instructor, and student-level fixed effects. Results from this estimate provide a number of intuitive results: students are less likely to drop courses if the course is in their major, if they have accumulated more credits, and if the course is taught by the same instructor who taught their prerequisite course. Students are more likely to withdraw if they are older and as time between prerequisite and follow-up courses grew. However, the coefficient associated with *Online*, representing the change in the probability of withdrawing from a follow-up given the prerequisite was completed online, is –0.0018 ($t = -0.28$, $p = 0.783$). Given the statistical insignificance of the impact of online prerequisites on withdrawing from future courses, it is unlikely that the prior estimates of β are biased because of nonrandom completion of follow-up courses.

5. DISCUSSION AND CONCLUSION

Students who take a prerequisite course in a traditional face-to-face setting perform better in subsequent courses than students who take a prerequisite course online. This difference is significant both statistically and practically, with the impact on follow-up grades of about one fourth of the difference between an A and an A– or, equivalently, about 0.22 standard deviations of the follow-up grade distribution. This effect is robust with respect to inclusion of student and course observables, as well as course, instructor, and time fixed effects. Broadly consistent with prior research, there is evidence that this negative impact of online instruction is larger for students who are weaker, as measured by performance in their prerequisite courses. The effect is slightly stronger when measured using the most immediate follow-up course than when measured as the average impact over all subsequent follow-ups.

One concern addressed by prior research has to do with nonrandom student selection into online courses. Whereas we present evidence that students taking online courses are younger and slightly less well academically prepared, when we instrument for online course enrollment using the physical distance between students' mailing addresses and campus, we determine that selection into online courses does not account for the finding that these courses cause reduced academic performance in follow-up courses.

A second type of selection bias could impact our findings. Because we examine academic performance in follow-up courses, one possible selection issue occurs if online courses deter (or encourage) students into enrolling into subsequent course work. For bias to occur, however, either academically weaker or stronger students must have differences in follow-up enrollment patterns. We show that there is no difference in general academic characteristics between students who do and do not enroll in follow-up courses. Moreover, we demonstrate that the completion rate of follow-up courses is statistically similar between students who took the online prerequisite and those who took a traditional prerequisite. Taken as a whole, it appears that selection into and out of follow-up courses is not responsible for the main finding that online courses lead to lower follow-up academic performance.

As pointed out in the Introduction, there are a number of studies that examine the impact of online courses. Ours differs from these in substantial ways. First, unlike the majority of studies, we explicitly account for nonrandom selection of students into online courses. Second, we approach this subject using all online courses taught at a university rather than focusing on a particular course. This latter point is important because it prevents drawing conclusions based on specific course experiences, which may or may not be more broadly generalizable to online education. Moreover, one expects institutions to create online policies impacting all courses at that institution rather than policies that affect single courses. Thus, the results found in this paper can better inform university administrators regarding their policies toward online instruction. Finally, unlike nearly all of the research on online courses, this paper explores the impact of taking a course online not on the performance in that course, but instead on the courses that follow-up the online course. This difference in methodologies is important as it remains unclear how online courses alter their grading and performance measures, thus potentially confounding performance measures results within online courses. Examining follow-up courses presents a more natural method of measuring the impact of online prerequisites as we do not expect instructors of follow-up courses to adjust their grading or evaluation methods based on past student experience in online courses.

In evaluating these results, it is important to remember that our analysis examines students who completed either an online or traditional prerequisite. This is important given that students in online courses tend to drop them prior to completion at about twice the rate of students in traditional courses, a fact consistent with the community college findings of Xu and Jaggars (2013). If weaker students are more likely to drop prerequisite courses (and in our case, never be in a position to enroll in a follow-up course), then an additional concern becomes apparent: Online courses would lead to slower progression through coursework as well as potentially lower performance in follow-up courses.

Although one may take these findings as an argument for eliminating online courses, we caution that our results speak only to the benefits, or lack thereof, of online teaching. It is certainly possible that the cost savings of providing online courses compensates for lower student performance. Nevertheless, although it may seem obvious that online courses are consistently cheaper than traditional ones, there is surprisingly little evidence on this subject. Most research on this topic is dated,[30] though Bacow et al. (2012) interviewed academic officers and reported that most believed online courses were at least as expensive as traditional courses due partly to substantive startup costs and partly to ongoing technological support needed for these courses. Whereas online courses may not reduce costs to the university, they potentially produce unmeasured benefits to the students taking them. For instance, student travel costs are lower and schedule flexibility is higher. Online courses potentially provide an alternative method of exploring material that students would not otherwise pursue, and online courses may reach audiences unable to attend traditional lectures. Although we find the provision of online courses leads to lower future achievement in related courses, we have not addressed the tradeoff between this lower future performance and these unmeasured benefits. We leave this question for future research.

ACKNOWLEDGMENTS

The authors wish to thank Beth Hartsoch for outstanding data assistance. Dan Hagen, David Hedrick, Dan Goldhaber, Roddy Theobald, Sharon Shewmake, Steven VanderStaay, and participants in the Economics Department Discussion Group at Western Washington University provided valuable comments that improved this paper.

REFERENCES

Allen, I. Elaine, and Jeff Seaman. 2013. *Changing course: Ten years of tracking online education in the United States*. Available www.onlinelearningsurvey.com/reports/changingcourse.pdf. Accessed 17 August 2015.

Bacow, Lawrence S., William G. Bowen, Kevin M. Guthrie, Kelly A. Lack, and Matthew P. Long. 2012. *Barriers to adoption of online learning systems in U.S. higher education*. New York: ITHAKA S+R. http://www.sr.ithaka.org/publications/barriers-to-adoption-of-online-learning-systems-in-u-s-higher-education/

Card, David. 1995. Using geographic variation in college proximity to estimate the return to schooling. In *Aspects of labour market behavior: Essays in honour of John Vanderkamp*, edited by Louis N. Christofides, E. Kenneth Grant, and Robert Swidinsky, pp. 201–222. Toronto: University of Toronto Press.

Crump, Richard K., V. Josesph Hotz, Guido W. Imbens, and Oscar A. Mitnik. 2009. Dealing with limited overlap in estimation of average treatment effects. *Biometrika* 96(1):187–199 doi:10.1093/biomet/asn055.

Figlio, David N., Mark Rush, and Lu Yin. 2010. Is it live or is it internet? Experimental estimates of the effects of online instruction on student learning. NBER Working Paper No. 16089.

Greene, William H. 2001. Estimating econometric models with fixed effects. Unpublished paper, New York University.

Hawkes, Mark, and Marge Cambre. 2000. The co$t factor: When is interactive distance technology justifiable? *T.H.E. Journal* 28(1):26–32.

Heckman, James J. 1981. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In *Structural analysis of discrete data with econometric applications*, edited by Charles F. Manski and Daniel McFadden, pp. 179–195. Cambridge, MA: MIT Press.

30. See Hawkes and Cambre (2000), Jewett (2000), Jung (2003), and Rumble (2003).

Imbens, Guido W. 2014. Matching methods in practice: Three examples. NBER Working Paper No. 19959.

Imbens, Guido W., and Donald B. Rubin. 2015. *Causal inference for statistics, social, and biomedical sciences*. New York: Cambridge University Press. doi:10.1017/CBO9781139025751.

Jaggars, Shanna Smith. 2014. Choosing between online and face-to-face courses: Community college student voices. *American Journal of Distance Education* 28(1):27–38.

Jaggars, Shanna Smith, and Thomas Bailey. 2010. *Effectiveness of fully online courses for college students: Response to a Department of Education meta-analysis*. New York: Community College Research Center, Columbia University.

Jewett, Frank I. 2000. BRIDGE: A model for comparing the costs of using distance instruction and classroom instruction. *American Journal of Distance Education* 14(2):37–47.

Jung, Insung. 2003. Cost-effectiveness of online education. In *Handbook of distance education*, edited by Michael Grahame Moore and William G. Anderson, pp. 717–726. Mahwah, NJ: Lawrence Erlbaum Associates.

Means, Barbara, Yukie Toyama, Robert Murphy, Marianne Bakia, and Karla Jones. 2010. *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development.

Parry, Marc, Kelly Field, and Beckie Supiano. 2013. The Gates effect. Available http://chronicle.com/article/The-Gates-Effect/140323./ Accessed 20 January 2014.

Rouse, Cecilia Elena. 1995. Democratization or diversion? The effect of community colleges on educational attainment. *Journal of Business & Economic Statistics* 13(2):217–224.

Rumble, Greville. 2003. Modeling the costs and economics of distance education. In *Handbook of distance education*, edited by Michael Grahame Moore and William G. Anderson, pp. 703–716. Mahwah, NJ: Lawrence Erlbaum Associates.

Schrager, Rick H. 1986. The impact of living group social climate on student academic performance. *Research in Higher Education* 25(3):265–276.

University of California Irvine (UCI). 2007. *Center of Assessment and Applied Research Commuter Report*. Irvine, CA: UCI.

Xu, Di, and Shanna Smith 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, Di, and Shanna 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):633–659.

Zhang, Dongsong. 2005. Interactive multimedia based e-learning: A study of effectiveness. *American Journal of Distance Education* 19(3):149–162.

Zhang, Dongsong, Lina Zhou, Robert O. Briggs, and Jay F. Nunamaker, Jr. 2006. Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information & Management* 43(1):15–27.