Online Learning, Offline Outcomes: Online Course Taking and High School Student Performance

Cassandra M. D. Hart

University of California, Davis

Dan Berger
Brian Jacob

University of Michigan

Susanna Loeb

Stanford University

Michael Hill iD

University of California, Davis

This article uses fixed effects models to estimate differences in contemporaneous and downstream academic outcomes for students who take courses virtually and face-to-face—both for initial attempts and for credit recovery. We find that while contemporaneous outcomes are positive for virtual students in both settings, downstream outcomes vary by attempt type. For first-time course takers, virtual course taking is associated with decreases in the likelihood of taking and passing follow-on courses and in graduation readiness (based on a proxy measure). For credit recovery students, virtual course taking is associated with an increased likelihood of taking and passing follow-on courses and being in line for graduation. Supplemental analyses suggest that selection on unobservables would have to be substantial to render these results null.

Keywords: *achievement, educational outcomes, high school, high school graduation, online courses, online instruction, virtual courses, virtual instruction*

Virtual education has expanded rapidly over the past 15 years. In 2002–2003, K–12 students took an estimated 317,000 virtual courses nationwide, as opposed to about 1.8 million enrollments in 2009–2010 (National Center for Education Statistics, 2012). New estimates suggest that this growth has continued apace in recent years, with an estimated 4.5 million K–12 course enrollments taken through online providers in 2014–2015 (Evergreen Education Group, 2015). Several states, including Florida, have established or are considering requirements for students to engage in at least one online learning experience before graduating, setting the stage for additional expansion of virtual course taking in coming years (Evergreen Education Group, 2015). Online learning advocates maintain that experience with online learning will be advantageous to students in an economy that rewards digital competency (Sheehy, 2012). Moreover, online course taking may reduce disparities in the quality of teaching across schools, as teachers are not tied to specific schools, and it may allow for pedagogical innovations.

Despite this steady growth in K–12 virtual course taking, we know little about how these courses affect student achievement (Means, Toyama, Murphy, & Bakia, 2013). On the one hand, proponents of virtual education point to several ways that virtual education could provide higher-quality education for students as compared with traditional classroom settings. For instance, virtual classes may allow students to work at a more individualized pace. This individualized pacing may help slower learners by allowing them to repeat confusing material until they master it, and it can help faster learners by allowing them to move on when they master material, without requiring them to sit through repetitious explanations (Berge & Clark, 2005; Tallent-Runnels et al., 2006). Virtual courses may also be well suited to provide immediate feedback on student performance to both students and teachers through intelligent tutoring systems, and they may provide for a uniquely interactive experience between students and the texts that they access (Means, Bakia, & Murphy, 2014). For instance, if students are able to click on links within lessons that provide them additional detail on a subject of interest, that will allow them to explore their interests interactively. Moreover, online courses allow students access to coursework, and potentially to high-quality teaching, that they may lack in their local school.

On the other hand, skeptics worry that online learning may be more difficult than learning in face-to-face environments. For instance, students who are inclined to procrastinate or who are not skilled in self-directed learning may suffer declines in performance if they lack a physically present teacher to direct their attention to the subject matter (Bork & Rucks-Ahidiana, 2013). Other students may be motivated but lack broadband or other technological resources that virtual courses rely on to enable smooth delivery. Even among students for whom technological resources do not pose a problem, some may have trouble in virtual courses if they lack the technological skills to make full use of the course content (Berge & Clark, 2005).

In this article, we examine virtual course taking in Florida—the state with the largest K–12 virtual sector in the nation—to determine the extent to which virtual course taking is associated with three types of outcomes: concurrent course performance, future course performance, and likelihood of persisting in high school through the final term of the 12th grade, which we use as a proxy for graduation. We distinguish first-time course takers from students who retake courses after failing on their initial attempt, and we further explore variation in virtual course-taking effects by student characteristics and course subjects. We focus on nine commonly taken academic courses (e.g., Geometry, English 1, World History) for which downstream course taking is common.

We find that among both first-time course takers and course retakers, those in virtual courses are more likely to pass the course (with a grade of C or higher) than are their face-to-face counterparts. These results are largely stable across different model specifications and for different subgroups of students and subject areas. The implications of this finding are not clear, however, since they could suggest that students learn more in virtual courses or that grading standards are more lenient in such courses.

When we look at downstream outcomes, we find a striking difference depending on whether a student is attempting the course for the first time or retaking it following an initial failure (credit recovery). For "first-time" course takers, virtual instruction is associated with moderately negative downstream outcomes. For example, first-time virtual course takers are 2.6 percentage points less likely to appear in public school records in the second term of their expected 12th-grade year, which we use as a proxy for expected high school graduation. Relative to the sample mean (77.8%), this represents a 3.3% decline in the likelihood of expected high school graduation. In contrast, we find that virtual course taking is positively associated with downstream outcomes for those students who were retaking the course. For example, virtual retakers are 6.5 percentage points more likely to appear in our data as a second-term 12th grader when compared with face-to-face students of the same course, even after we control for an extensive set of student

and school characteristics. This translates to a 10% higher likelihood of expected graduation.

Our results are robust to alternative model specifications. Moreover, both the contemporaneous course outcomes and the downstream outcomes appear stable to assumptions about selection. Using methods developed by Oster (2017), we find that once we control for a rich set of observable student and school characteristics, the selection on unobservable characteristics would have to be extremely large to invalidate our main results.

What might explain the differences that we find in online course taking by attempt type (first-time or retake)? One possibility is that these differences could stem from differences in the characteristics of the students in the two types of classes. However, since we find comparable results across a variety of measured student characteristics, this explanation seems less likely. Another possibility is that the differences arise from differences in the counterfactual course offerings. For example, traditional options for students to retake a course in a face-to-face setting, as in summer school, may be lower quality in terms of instruction as compared with the standard version of the course that students take as a first attempt. Students may also not be motivated to work hard during the summer, when traditional retake opportunities are common. Alternatively, perhaps something about course retaking makes it more difficult to accomplish in a traditional face-to-face setting. For instance, students may be embarrassed to repeat the class in view of their peers (whether the course is retaken during the academic year with younger children or at summer school). Although we cannot test these mechanisms directly, our findings of differences in outcomes for first-time versus credit recovery attempts can help guide thinking through how best to use virtual courses for student benefit.

Background

While the use of virtual courses among K–12 students has grown rapidly, credible research on the effects of those classes on student performance remains scant (Barbour, 2013; Means et al., 2013). A larger body of literature has addressed the relationship between virtual course taking and academic performance in higher education settings. One set of studies that randomized students to online/hybrid or face-to-face versions of classes tended to find null to negative effects of virtual instruction (Bowen, Chingos, Lack, & Nygren, 2014; Figlio, Rush, & Yin, 2013; Joyce, Crockett, Jaeger, Altindag, & O'Connell, 2015). These studies, while internally valid, relied on samples from relatively selective institutions that often enroll unusually motivated and well-prepared student bodies. The generalizability of these results is unclear.

A second set of studies in higher education used quasi-experimental methods to explore the effects of virtual course

taking, largely among broad access institutions such as community colleges or for-profit institutions. Studies conducted in Virginia (Xu & Jaggers, 2011), Washington (Xu & Jaggers, 2013), California (Hart, Friedmann, & Hill, 2018), and anonymized state systems (Streich, 2014) consistently found poorer performance for students who take classes virtually. Recent evidence suggests that online course taking is associated with negative downstream impacts on performance in follow-on courses as well (Krieg & Henson, 2016). Looking at a different type of broad access setting—a large for-profit university—Bettinger, Fox, Loeb, and Taylor (2017) found similar patterns, with students in online courses performing less well than their peers contemporaneously and facing reduced likelihood of continuing enrollment.[1]

These same relationships may be at play in the K–12 sector, or plausibly, the effects of online course taking may differ. While college instructors typically have broad latitude to set the terms of their own courses—determining course content, assignment structures, and so on—such latitude may be considerably reduced at the secondary level, where state standards place stricter limits on what content must be covered in courses and where organizational practices may require all teachers served by the virtual school to use a proscribed set of instructional materials (Friend & Johnson, 2005). Moreover, secondary students are in school full-time and likely have fewer responsibilities, such as work or family, which may crowd out time for school work.

Two studies at the high school level used randomized controlled trials to evaluate the learning effects of virtual or hybrid course taking. The first study randomized ninth-grade students taking Algebra 1 to face-to-face or hybrid models of course delivery (Cavalluzzo, Lowther, Mokher, & Fan, 2012). Hybrid versions of courses offered teachers access to customizable courseware provided by Kentucky Virtual Schools and the Kentucky Department of Education. Roughly 60% of instruction was delivered face-to-face, with 40% delivered through the virtual courseware. Researchers found no differences on test scores between groups randomized to the traditional and hybrid versions of the course, although treatment fidelity was an issue in the study (Cavalluzzo et al., 2012).

A second randomized controlled trial explored effects on student outcomes in Algebra 1 credit recovery classes offered over the summer months (Heppen et al., 2017). Researchers randomly assigned students who had failed second-semester Algebra 1 to face-to-face or online credit recovery courses (both of which took place in the summer). Students assigned to the online delivery mode rated the class as being more difficult, and credit recovery success rates and algebra posttest scores were both higher in the face-to-face condition. Longer-term outcomes for the first cohort of the study, however, suggested no lasting

differences between online and face-to-face students on performance in downstream courses.

A handful of quasi-experimental studies came to mixed conclusions about the educational effects of virtual education among students who opt into those settings. A recent study by the Center for Research on Education Outcomes compared students in virtual charter schools with "virtual control records"—that is, records for students who look observationally similar to virtual charter students. The center found that students in virtual charters made substantially smaller gains on math and reading scores over a year than did their peers who remained in face-to-face schools (Woodworth et al., 2015). A study specific to fully online schools in Ohio found similarly negative effects of enrollment in virtual schools (Ahn & McEachin, 2017). Likewise, a quasi-experimental study examining virtual course taking in algebra found that students induced into virtual classrooms underperformed their peers in face-to-face environments (Heissel, 2016).

The study closest to our own is a working paper by Chingos and Schwerdt (2014). Their study examined how participation in Florida Virtual School (FLVS) courses affected student scores on high school math and English standardized tests. They found that although students who took FLVS courses constituted a positively selected group relative to their peers, even when controlling for student and school characteristics, students who took FLVS courses had null or slightly positive results as compared with their peers on standardized tests.

Our article extends on theirs and others examining online course taking in K–12 in several ways. First, we explore a novel set of outcomes for the K–12 level: concurrent course performance, performance in downstream courses, and progression through school. Second, we explore a broader range of courses. Our main results assess virtual course taking in a range of commonly taken courses in English, math, science and social studies, and in robustness checks, we broaden our course sample to look at concurrent course performance in a broader range of courses. Most previous studies focused primarily on math and English language arts. Third, we explore and distinguish the performance outcomes of first-time course takers from those retaking courses to recover credit, which turns out to be informative. We might expect the effects to differ between first-time takers and retakers because the latter are likely to need access to courses at nontraditional times and may retake their courses while fulfilling other traditional requirements. Thus, the counterfactual for online course taking may differ for the two groups. The credit recovery group is also important to study because a growing number of districts, particularly in Florida, are relying on online options to provide credit recovery instruction to supplement and supplant traditional options, such as summer school (Gonzalez, 2012).

Setting

We study online course taking in the context of Florida's virtual education system. Florida is home to the most extensive K–12 virtual education system in the country, and the Florida education code requires students entering ninth grade since the 2011–2012 school year to take at least one course online. To facilitate this graduation requirement, the state also requires that each district offer every K–12 study multiple part- and full-time options for virtual course taking (Florida Statute 1002.321, 2018). Two such options come through part- and full-time programs operated by FLVS: the statewide virtual school freely available to public school students who meet a broad set of eligibility criteria.[2] FLVS is the largest K–12 virtual education provider in the nation, serving students through >400,000 enrollments in 2014–2015 (Evergreen Education Group, 2015).[3] Districts may operate their own virtual schools or district-based franchises of FLVS. Franchises use FLVS-developed courses and have access to FLVS-provided professional development, but content is delivered by district-employed teachers (Evergreen Education Group, 2015). In addition, other virtual providers increasingly serve Florida students, including K12, Connections Education, and others.

Our data include many virtual providers, and students participate in a range of virtual courses, which can vary in their materials, teachers, type and extent of interactions, and other features. FLVS is the predominant virtual provider in our data. FLVS courses are generally asynchronous; in other words, students may all be accessing course materials at different times. However, instructors are required to have regular phone check-ins with students to assess their understanding and address their questions. In this way, students have regular real-time instructor interactions as well as online asynchronous interactions (Jacob, Berger, Hart, & Loeb, 2016). FLVS teachers work with about 150 students on average, roughly equivalent to a high school teacher with six periods of 25 students each (FLVS, n.d.).

Methods

Data and Sample

We draw on data from the Florida Education Data Warehouse, maintained by the Florida Department of Education. Data include course enrollments taken through traditional public schools as well as credits received from online instructional institutions. Information on some additional school characteristics comes from the National Center for Education Statistics and from Florida School Indicator Reports.

Our main analytic sample includes students taking courses in Grades 9 and 10 in Florida high schools for the 2006–2007 through 2011–2012 school years. Throughout the article, we designate school years by referring to the calendar year of the spring term (e.g., 2012 for 2011–2012). We use data as early as 2005 to characterize students' prior performance on standardized tests in middle school and to determine whether students had previously taken courses, and we use data as late as 2014 to determine whether students take follow-on courses. We focus on ninth and 10th graders so that we can assess effects on future coursework and because course selections vary less across students in those grades. In robustness checks, we expand our sample to include students in 11th and 12th grades.

Because one of our key outcomes of interest is performance in subsequent courses, we focus on a subset of courses that are commonly taken, are considered academic courses (as opposed to life skills courses, such as physical education or drivers' education), and are likely to be taken in a specific sequence. We are able to identify a set of courses in which the majority of course takers subsequently appear within the next 2 years. To ensure that we have a consistent 2-year look-forward period in which to observe future course taking for all students, we focus on ninth- and 10th-grade students, and our target courses therefore include courses primarily taken in those grades—specifically, Algebra 1, Biology 1, Chemistry, Earth/Space Science, English 1, English 2, Geometry, Physical Science, and World History.

While these courses have advantages in that we are able to look at downstream course outcomes, these are not the courses most frequently taken online. As Table 1 shows, students take only about 1.3% of their first-attempt enrollments in "core" academic subjects, such as math, English language arts, social studies, and sciences online, while they take about 4.6% of enrollments virtually in life skills courses, such as physical education and driver's education. Online enrollment is similarly less prevalent for credit recovery in these core academic subjects than for other types of courses. Moreover, our need to look forward for 2 years to capture outcomes in future courses limits us from examining outcomes in 11th and 12th grades, which are the grades where online course taking is more prevalent in Florida (Jacob et al., 2016). By using robustness checks, we address the external validity concerns associated with our sample limitations in terms of courses and grade levels.

Because our analytic strategies, detailed later, rely on comparing virtual course takers' outcomes with outcomes of peers in the same school, our study focuses on students enrolled in brick-and-mortar schools. We therefore exclude students enrolled in virtual schools full-time. These students, who take virtual classes to supplement instruction provided at their brick-and-mortar schools, generally take a relatively small share of courses online. Among students in our sample who took any course virtually in at least one of nine target courses, >80% of them took only one course virtually.

We also distinguish between types of course attempts: first-time attempts and retake efforts. Our first-attempt sample comprises all ninth- and 10th-grade students in target

TABLE 1
Share of Enrollments Taken Virtually

	All students, %
All courses	2.023
Courses taken for the first time (i.e., first attempt)	1.773
Core subjects	1.315
Life skills	4.641
Foreign languages	3.218
Other electives	1.213
Observations, *n*	44,726,942
Courses taken after prior failure (i.e., credit recovery)	6.185
Core subjects	5.903
Life skills	6.695
Foreign languages	14.107
Other electives	4.665
Observations, *n*	909,887
Observations, *n*	45,636,829

Note. Tabulations are based on Florida Department of Education data. All values are presented as percentages unless noted otherwise. Includes years 2007 to 2014 (by spring semester) and summer courses. Core courses include math, English, language arts, science, and social studies and exclude Advance Placement / International Baccalaureate courses. Life skills include physical education, driver's education, and health.

courses who were not previously enrolled in a given course. Our course-retaking sample includes students who failed one of the target courses as ninth or 10th graders. Failure is defined by having a cumulative-course grade point average <1.0 across semesters, which equates to less than a D. A grade point performance <1.0 implies that a student failed at least one semester of the course. Our sample includes students who retook the course within the next academic year (including the summer after the academic year of initial failure; roughly 15% of retake enrollments occur during the summer). We exclude students who did not retake courses the next summer or academic year.[4] In robustness checks, we expand our sample to include course retake attempts within 2 years of failing the initial course.

Measures

Our main variables of interest are indicators for whether a class is taken through a virtual provider (vs. an omitted category of exclusively taken face-to-face). Course mode is derived per the instructional institution that the high school transcript files record for the course enrollment. Note that for credit recovery courses, course mode refers to the mode of the retake effort, not the mode of the initial course. In a small number of cases (<3%), instructional institution records are missing, but data are available on the enrollment institution (the institution where the student is enrolled). In these cases, we assume that instruction is provided in the enrollment institution; we show in robustness checks that our results are insensitive to the exclusion of these records.

Outcomes. We explore three main outcomes of interest. Our first outcome of interest is concurrent course performance. Because grades are ordinal, we focus on a binary indicator for whether the student passes the class with a grade of C or better. This measure has substantive importance for students because students need a final grade point average ≥2.0 to graduate (Florida Department of Education, 2016a).

A concern with using contemporaneous course performance as an outcome measure is that grading standards may vary across teachers and schools and between the online and in-person modality. To address this concern, we look at enrollment and performance in follow-on courses (Carrell & West, 2010; Figlio, Schapiro, & Soter, 2015). Follow-on courses are defined as the next course taken in the same subject area.[5] Our main outcome of interest for follow-on coursework is a measure that captures whether a student takes and passes a follow-on course with a grade of C or better. This outcome captures whether a student goes on to any successful future course taking in the same subject area. In supplemental analyses, we decompose the future course-taking behavior by looking at (1) the decision of whether to enroll in follow-on courses in the same subject, (2) the characteristics of the follow-on courses taken, and (3) the likelihood that students pass follow-on courses, conditional on taking one and controlling for next-course characteristics; this decomposition allows us to explore the mechanisms through which virtual course taking may affect future course outcomes.

A third outcome measures the likelihood that a student is observed in the second semester of his or her projected senior year. That is, we look at whether we observe students earning credits in the final term of their projected 12th-grade year or if we see them earning credit for a full-year class (implying that they were attending in the final term). We project that current ninth and 10th graders should be in 12th grade in 3 and 2 years, respectively. We lack graduation data and so cannot observe whether a student actually earns a degree, but we regard this as a strong proxy for students being on track to graduate. In supplemental analyses, we look at credits earned as an alternate measure and find similar results.[6]

Student Controls. To compare observably similar students, we include a host of student controls. A vector of race indicators indexes whether students are White (omitted), Black, Hispanic, Asian, multiracial, or "other race"; gender is captured by an indicator for whether students are female.

Student socioeconomic status is captured by an indicator recording whether a student is eligible for free or reduced-price lunch (FRPL) in a given year.

We include several variables that capture students' prior academic performance. We capture prior test performance through grade-by-year standardized student scores on the eighth-grade math and English language arts sections of the Florida Comprehensive Assessment Test (FCAT), which Florida used for accountability purposes. We also include indicators for whether a student is enrolled in a gifted program, identified for special education programs, or classified as limited English proficient in a given year. We include the student's attendance rate in the same year as the course is taken, as an additional control. We also include indicators for whether each student was a member of a cohort that was subject to an online course-taking requirement to graduate. Students entering ninth grade in fall 2011 or later were required to take at least one online course prior to graduation.[7]

Our most saturated models include controls for ninth-grade academic measures (when observed for 10th-grade students) and a more complete set of middle school controls. Ninth-grade academic measures include FCAT scores, attendance rate, Grade 9 grade point average, and number of credits earned. More saturated middle school controls include sixth- to eighth-grade attendance rates, sixth- and seventh-grade FCAT scores in math and reading, and quadratics for test terms and interactions between the math and reading scores for each grade level.

Home Institution Characteristics. We control for a series of characteristics of students' home institutions. Home institutions refer to the brick-and-mortar school that students attend, while the instructional institution refers to the institution providing a specific course. School academic quality is captured by a series of indicators for the grade (A–F) received by the school under the state's accountability plan in the current academic year. As a second measure of student achievement, we include the mean value of the eighth-grade FCAT scores of the incoming cohort of ninth graders. School demographic measures include the share of the student body that is Black, Hispanic, Asian, or other race (percentage White is omitted) and the share of students using subsidized lunch. We also include indicators for whether the home institution is a charter or magnet school, as well as a series of indicators capturing school urbanicity (city and suburb vs. rural and town).

Analytic Methods

In assessing the relationship between virtual course taking and student outcomes, we are concerned about several types of selection bias. First, courses disproportionately taken online may be harder or easier to pass than courses taken less frequently online. Table 1 shows the prevalence of virtual course taking among different class types. Note that many of the most popular classes are in life skills (e.g., physical education, driver's education) or subject areas (e.g., foreign languages), while online course taking is less common in core courses (e.g., English and math). The difference in courses taken online and in-person may bias our estimates if pass rates are systematically higher or lower in courses that enroll larger shares of students virtually. To address this concern, we include course fixed effects. Year fixed effects and grade fixed effects similarly adjust for potential differences in course taking across time and grades.

To examine the relationship between virtual course taking and academic performance, we estimate ordinary least squares regression of outcome Y for student i in course c in grade g in school s in year t:

$$Y_{icgst} = \beta\text{Online}_{icst} + \theta_c + \theta_t + \theta_g + \varepsilon_{icgst}, \qquad (1)$$

where θ_c is a course fixed effect (indicating whether the class is Algebra 1, English 1, etc.), θ_t is a year fixed effect, θ_g is a grade fixed effect, *Online* is an indicator for whether a course is taken online, and ε_{icgst} is an independently and identically distributed error term.

These simple regressions may be subject to bias from school and student factors. For instance, descriptive statistics (Table 2) suggest that virtual students were more likely than face-to-face students to come from schools that received state report card grades of A and B. If students in these relatively advantaged schools benefit from other resources that promote success in future course taking, our results could be positively biased. With respect to student factors, virtual classes enroll a higher share of female students and gifted students and a lower share of students eligible for FRPL, English learners (students designated as limited English proficient), and students using special education. Virtual course takers also have, on average, higher eighth-grade math and English language arts scores on the FCAT. These student factors are independently associated with virtual course taking (in the same direction) if we model virtual course-taking behaviors including all characteristics in the same regression, even when we control for school attended (results available on request). If more academically advantaged students disproportionately enroll in virtual courses, as these descriptive statistics suggest, estimates could be biased upward, unless the models adequately adjust for incoming differences.

We address these concerns by estimating highly saturated fixed effects models, including student and school controls, as well as school fixed effects θ_s:

$$Y_{icgst} = \beta\text{Online}_{icst} + \delta\text{StudentChar}_{it} + \\ \lambda\text{SchoolChar}_{st} + \theta_s + \theta_c + \theta_t + \theta_g + \varepsilon_{icgst}, \qquad (2)$$

TABLE 2
Summary Statistics

	First attempt sample		Credit recovery sample	
	Took face-to-face	Took virtually	Retook face-to-face	Retook virtually
Outcomes				
Passed concurrent course (C or better)	0.713	0.849	0.579	0.683
Took and passed next course	0.653	0.642	0.415	0.479
Persist to expected second semester, 12th-grade year	0.778	0.728	0.650	0.769
Student characteristics				
Female	0.492	0.536	0.398	0.457
White	0.442	0.612	0.342	0.519
Black	0.221	0.149	0.309	0.200
Hispanic	0.284	0.168	0.315	0.229
Asian	0.023	0.031	0.010	0.015
Multiracial	0.024	0.033	0.019	0.030
Other race	0.006	0.008	0.005	0.007
Free or reduced-price lunch	0.480	0.360	0.613	0.445
Gifted	0.048	0.070	0.017	0.037
Limited English proficiency	0.202	0.086	0.242	0.120
Special education	0.113	0.093	0.163	0.113
FCAT eighth-grade score (standardized)				
Math	0.025	0.180	−0.355	−0.088
Reading	0.012	0.195	−0.339	−0.065
Grade 9	0.495	0.475	0.548	0.514
Grade 10	0.505	0.525	0.452	0.486
School characteristics				
Percentage White	48.167	56.568	43.083	52.161
Percentage Black	22.430	18.704	25.638	19.953
Percentage Hispanic/Latino	25.491	19.966	27.760	23.373
Percentage Asian	2.583	2.918	2.413	2.940
Percentage other race	1.215	1.812	0.940	1.539
Percentage free or reduced-price lunch	41.436	37.939	43.864	38.619
Urban	0.254	0.254	0.276	0.261
Suburban	0.592	0.552	0.592	0.608
Rural	0.150	0.192	0.127	0.129
Charter	0.031	0.029	0.022	0.032
Magnet	0.306	0.252	0.338	0.268
Average FCAT score of incoming ninth graders (standardized)				
Math	0.037	0.132	−0.039	0.113
Reading	0.033	0.128	−0.044	0.103
School accountability grade				
A	0.274	0.361	0.201	0.335
B	0.327	0.375	0.287	0.367
C	0.225	0.155	0.264	0.180
D	0.142	0.093	0.194	0.107
F	0.027	0.007	0.050	0.009

TABLE 2 (CONTINUED)

	First attempt sample		Credit recovery sample	
	Took face-to-face	Took virtually	Retook face-to-face	Retook virtually
Subject area				
Math	0.278	0.308	0.441	0.351
Science	0.313	0.278	0.219	0.250
English	0.229	0.228	0.223	0.241
Social studies	0.180	0.187	0.117	0.159
Proportion of sample	0.991	0.009	0.918	0.082

Note. Tabulations are based on Florida Department of Education data. Binary values are presented as proportions unless noted otherwise. Samples include ninth- and 10th-grade student enrollments for the 2007–2012 academic years (by spring semester) in the following courses: Algebra 1, Biology 1, Chemistry, Earth/Space Science, English 1, English 2, Geometry, Physical Science, and World History. Credit recovery students all failed a prior attempt at the class (earned a course average of <1.0). FCAT = Florida Comprehensive Achievement Test.

SchoolChar is a vector of characteristics of the home institution, including racial composition, share of students on FRPL, the average eighth-grade math and English language arts scores of incoming ninth graders, urbanicity of the school, school accountability grade, and charter or magnet status. *StudentChar* is a vector of student characteristics, including sex, race, FRPL use, gifted status, limited English proficiency designation, special education status, attendance rate, an indicator for whether the student was subject to a state online course graduation requirement, ninth-grade academic performance indicators, and middle school test and attendance measures. All models include indicators for missing variables. We use robust standard errors clustered at the school level.

These results may still be subject to bias if students who take virtual courses differ in unobservable ways—such as self-motivation—from their same-school peers. This concern may be especially pronounced because our results are based on a relatively modest proportion of students who take courses online. To address this concern, we provide evidence on the degree of selection on observables required to render our school fixed effects models null, relying on procedures detailed in Oster (2017). As we describe in greater detail, we find that our results would require a substantial degree of selection on unobservables to render our main results null.

Results

Main Results

First-time course taking. Table 3 presents the main results for our first-time course-taking sample. The five columns present model estimates, with each subsequent model adding controls—from column 1, which includes only course, year, and grade fixed effects, through column 5, which includes a host of student and school covariates, including student performance measures from middle school. The effects are mixed. We find positive effects on passing the contemporaneous course but negative effects on subsequent course performance as well as our graduation proxy. For example, the specification in column 3, which contains basic student and school controls, suggests that taking a course virtually increases the likelihood of passing it by 12.5 percentage points (roughly 18%). However, it reduces the likelihood of taking and passing a follow-on course in the same subject by 1.5 percentage points (roughly 2%) and lowers our proxy for high school graduation by 3.4 percentage points (4%).

Because we remain concerned about bias, we add student-level controls to check for specification errors. In particular, we first add controls for ninth-grade performance, which are set to 0 for ninth-grade students, with grade fixed effects subsuming the missing variable indicators. Our downstream coefficients decline in magnitude across specifications—falling in magnitude by nearly half for the next-course outcome—but remain significant (column 4). As a final set of controls, we add in the richer set of middle school controls, including attendance rates and lagged test scores back to sixth grade, in addition to quadratic terms and interaction terms for the math and reading scores in each year. These added controls make little additional difference, although the significance of the next-course estimates diminishes to $p < .10$ (column 5); the relative stability of the estimates with these saturated sets of controls give us some confidence that our models likely account for a high degree of selection with the controls in place.

Course retakes. We find a somewhat different pattern in the effects of virtual course taking on course retake attempts (Table 4). When our full set of controls and school fixed effects is included (column 5), students who repeat virtual courses are 4.7 percentage points more likely to pass their remedial course, 1.7 percentage points more likely to jointly take and pass future same-subject courses, and 6.5 percentage points more likely to be observed in a projected final term in senior year, as compared with peers who retake coursework in face-to-face settings ($p < .01$).

TABLE 3
First-Time Course Takers: Main Outcomes

	1	2	3	4	5
		A: Course passing			
Took course virtually	0.139*** (0.005)	0.116*** (0.004)	0.125*** (0.005)	0.131*** (0.005)	0.131*** (0.005)
Comparison group mean	0.713				
Proportion taking course virtually	0.009				
Observations, n	5,941,424	5,941,424	5,941,424	5,941,424	5,941,424
R^2	0.019	0.172	0.189	0.22	0.222
		B: Took and passed next course			
Took course virtually	−0.007 (0.005)	−0.021*** (0.004)	−0.015*** (0.004)	−0.008** (0.004)	−0.007* (0.004)
Comparison group mean	0.653				
Proportion taking course virtually	0.009				
Observations, n	5,961,729	5,961,729	5,961,729	5,961,729	5,961,729
R^2	0.019	0.139	0.152	0.171	0.173
		C: Graduation proxy (final term, 12th-grade year)			
Took course virtually	−0.054*** (0.006)	−0.037*** (0.003)	−0.034*** (0.003)	−0.027*** (0.003)	−0.026*** (0.003)
Comparison group mean	0.778				
Proportion taking course virtually	0.009				
Observations, n	5,429,093	5,429,093	5,429,093	5,429,093	5,429,093
R^2	0.016	0.161	0.173	0.186	0.191
		Model specifications			
Initial student controls		X	X	X	X
School controls, fixed effects			X	X	X
Grade 9 controls				X	X
Expanded Grade 6–8 controls					X

Note. Tabulations are based on Florida Department of Education data. Standard errors (in parentheses) are clustered at the school level. All models include fixed effects for course, academic year, and grade level. Initial student-level controls include gender, race, FRPL, gifted status, limited English proficiency status, attendance rate, online requirement graduation cohort indicator, and eighth-grade math and English language arts scores. School-level controls include racial composition; percent free lunch, eighth-grade math and English language arts scores of incoming ninth graders, urbanicity, school accountability grade, and charter or magnet status. Grade 9 controls include Florida Comprehensive Achievement Test scores, attendance rate, grade point average, and number of credits earned in Grade 9. Grade 6–8 controls include expanded test score controls (quadratic and interaction terms) and attendance rates. Missing variable dummies are included in all models.
*$p < .1$. **$p < .05$. ***$p < .01$.

We run a variety of robustness tests to verify that our results are not an artifact of sample selections and analytic specifications (see online Appendix B). We find that the pattern and significance of our results are, for the most part, robust to different sample and specification decisions, although the magnitude of the coefficients shifts somewhat depending on these decisions. Specifically, we confirm that our positive concurrent course-passing results for first-time attempts and retakes hold—albeit with a notable reduction in magnitude for first-time course takers—when we expand our sample to include a broader range of grades and courses. We confirm that results are insensitive to excluding the 1% to 2% of records for which we made assumptions that the student's home institution provided instruction in the absence of explicit records of instructional institution.

We also confirm that our results are generally insensitive to specification decisions. Our results are similar when we experiment with the inclusion of different patterns of fixed effects (including school by course, course by grade by year, and school by grade by year). We confirm that we find similar point estimates if we use a propensity score–matching analytic strategy. We further confirm that our remedial course-taking results are similar if we include courses that students took within 2 years of the initial failure; our main models include only courses retaken within 1 year. We also confirm that we see a similar set of results to our main graduation proxy if we use alternative (but potentially less reliable) measures of credits accumulated by a projected 12th-grade year. We discuss these results in detail in the online Appendix B.

TABLE 4
Course Retakers: Main Outcomes

	1	2	3	4	5
			A: Course passing		
Retook course virtually	0.083*** (0.011)	0.053*** (0.011)	0.056*** (0.011)	0.049*** (0.011)	0.047*** (0.011)
Mean of outcome	0.579				
Proportion taking course virtually	0.082				
Observations, n	322,546	322,546	322,546	322,546	322,546
R^2	0.023	0.055	0.110	0.115	0.116
			B: Take and pass next course		
Retook course virtually	0.051*** (0.005)	0.027*** (0.005)	0.027*** (0.004)	0.020*** (0.004)	0.017*** (0.004)
Mean of outcome	0.415				
Proportion taking course virtually	0.082				
Observations, n	329,034	329,034	329,034	329,034	329,034
R^2	0.012	0.037	0.053	0.058	0.059
			C: Graduation proxy (final term, 12th-grade year)		
Retook course virtually	0.107*** (0.005)	0.079*** (0.004)	0.076*** (0.004)	0.068*** (0.004)	0.065*** (0.004)
Mean of outcome	0.650				
Proportion taking course virtually	0.082				
Observations, n	306,257	306,257	306,257	306,257	306,257
R^2	0.051	0.136	0.153	0.162	0.166
			Model specifications		
Initial student controls		X	X	X	X
School controls, fixed effects			X	X	X
Grade 9 controls				X	X
Expanded Grade 6–8 controls					X

Note. Tabulations are based on Florida Department of Education data. Standard errors (in parentheses) are clustered at the school level. Regressions are modified as described in panel titles. All models include fixed effects for course, academic year, and grade level. Initial student-level controls include gender, race, free lunch, gifted status, limited English proficiency status, attendance rate, online requirement graduation cohort indicator, and eighth-grade math and English language arts. School-level controls include racial composition, percentage free lunch, eighth-grade math and English language arts scores of incoming ninth graders, urbanicity, school accountability grade, and charter or magnet status. Grade 9 controls include Florida Comprehensive Achievement Test scores, attendance rate, grade point average, and number of credits earned in Grade 9. Grade 6–8 controls include expanded test score controls (quadratic and interaction terms) and attendance rates. Missing variable dummies are included in all models.
***$p < .01$.

Selection

Notwithstanding the general stability of the results established in our robustness tests, we may still be concerned that the results are an artifact of selection. Perhaps especially motivated or tech-savvy students opt into virtual courses. While we cannot conclusively speak to how seriously selection issues are affecting our results, we can speak to the extent to which selection would have to occur to render our results null. Building on the work of Altonji, Elder, and Taber (2005), Oster's procedure (2017) uses the change in magnitude of "treatment" coefficients when controls are included, as compared with uncontrolled models, to quantify the extent to which the inclusion of observable characteristics reduces bias and the extent to which additional selection on unobservables would have to exist to render the "treatment" effect null. Specifically, Oster's procedure determines

the degree of selection on unobservables (δ) necessary to return a null coefficient, given a specified R^2. The specified R^2 gives the degree of variance that researchers would expect that a model including observable and unobservable covariates would reasonably explain. Oster suggested that a δ of 1 represents a relatively stable result, as it suggests that selection on unobservables would have to be more severe than selection on observables to overturn the result (i.e., to return a coefficient, β^*, of 0).

A key decision in bounds analysis involves the choice of a reasonable maximum R^2 (R_{max}). Oster (2017) found that an R_{max} of $1.3 \times R^2$ from a preferred model weeds out a high proportion of unstable coefficients while retaining most significant findings from randomized controlled trials.[8] In our case, our preferred models use school fixed effects, which should control well for unobservable school factors. The

TABLE 5

Bounds Analysis Accounting for Selection on Unobservable Characteristics

	First-time course takers			Credit recovery		
	1: Pass concurrent course	2: Take/pass follow-on course	3: Graduation proxy	4: Pass concurrent course	5: Take/pass follow-on course	6: Graduation proxy
A: R^2 statistics						
School fixed effects models						
All controls (Model 5)	0.222	0.173	0.191	0.116	0.059	0.166
No student controls	0.046	0.043	0.037	0.083	0.031	0.078
B: Bounds analysis given R_{max} = R^2 inflated by 1.3 × student controls contribution						
δ for β^* = 0	80.905	7.800	4.324	7.685	2.498	6.26
Estimated R_{max}	0.274	0.212	0.237	0.126	0.067	0.192

Note. Tabulations are based on Florida Department of Education data. Deltas are calculated with psacalc, which implements Oster's (2017) procedure for determining the degree of proportional selection on unobservables required to return a 0 coefficient. We assume that R_{max} is bounded by an adjustment factor of 1.3 applied only to the portion of the R^2 affected by student controls (assuming that unobservables are driven by student, rather than school, selection).

remaining selection likely comes from differences in the types of students who select into virtual courses, rather than from cross-school differences. We therefore estimate coefficient stability assuming that the adjustment factor of 1.3 applies only to the marginal explanatory power of the student observable characteristics (i.e., to the improvement of R^2 that comes after the addition of student controls to models that already include all school controls and fixed effects).[9] Table 5, panel A, provides the R^2 statistics from school fixed effect models that use all controls (student and school, equating to Model 5 in Tables 3 and 4), as well as a version of the school fixed effect models omitting student controls (but including all other controls). In panel B, for each value of R_{max}, we show the δ required to return a coefficient of β^* = 0. The value of R_{max} used to generate estimates is given in the final row for each outcome.

Results suggest that our coefficients are fairly robust to selection. For all results, we see δ > 1, in most cases by a large factor. These results provide evidence that our results are reasonably stable and that selection on unobservables would have to be quite large to overturn our main results.[10]

Decomposition of mechanisms in future course taking. Virtual course taking may affect the joint likelihood of taking and passing future coursework in the same subject in multiple ways: by changing the likelihood of taking a follow-on course, by changing the characteristics (e.g., course difficulty or instructional model) of the next course attempted, or by changing actual performance based on prior learning. In the online Appendix C, we separate these channels.

We find that our positive results for jointly taking and passing the next course for credit recovery students are driven largely by the likelihood of taking a follow-on course. Credit recovery students are 4 percentage points more likely to take a follow-on course, and students who take courses

virtually for both attempt types are more likely to select into courses that we would expect to have higher pass rates, due to an increased likelihood of taking follow-on courses virtually and general course difficulty proxied by overall pass rates. Conditional on these course characteristics and controlling for next-course fixed effects, virtual students are effectively no more or less likely to pass their follow-on courses than are face-to-face students for first-time course taking, while they are modestly more likely to pass follow-on courses after credit recovery efforts.

Heterogeneity of Results by Student Characteristics

Our main pattern of results describes average effects, but different types of students and different types of courses could see different effects of virtual course taking. The positive association between virtual course taking and contemporaneous course performance for first-time course taking is largely consistent across students with different background characteristics, although the size of the coefficient varies somewhat, particularly for students of different prior achievement profiles (Table 6). Negative associations between virtual course taking and being positioned for graduation are likewise consistently negative across most student subgroups. By contrast, the negative association between virtual course taking and follow-on course outcomes is largely driven by more advantaged students (White, non–FRPL using, higher achieving). Lower-achieving students (i.e., those in the bottom quartile of a measure of average standardized eighth-grade FCAT scores) have a more positive, or at least less negative, pattern of results in virtual first-attempt classes than do higher-achieving students for all outcomes.

The subgroup patterns look quite different for course retakers: for remedial course taking, more advantaged students

TABLE 6

Heterogeneity by Student Subgroup

	First-time course takers			Credit recovery		
	1: Pass concurrent course	2: Take/pass follow-on course	3: Graduation proxy	4: Pass concurrent course	5: Take/pass follow-on course	6: Graduation proxy
White						
Took course virtually	0.136*** (0.005)	−0.009** (0.004)	−0.022*** (0.004)	0.090*** (0.011)	0.024*** (0.005)	0.078*** (0.005)
Mean of outcome	0.74	0.67	0.76	0.55	0.39	0.58
Proportion taking virtually	0.012	0.012	0.012	0.120	0.120	0.012
Observations, *n*	2,635,215	2,641,311	2,427,091	115,126	117,337	109,529
Black						
Took course virtually	0.156*** (0.008)	−0.005 (0.008)	−0.037*** (0.007)	0.015 (0.015)	−0.006 (0.009)	0.056*** (0.007)
Mean of outcome	0.64	0.60	0.78	0.58	0.42	0.68
Proportion taking virtually	0.006	0.006	0.006	0.055	0.055	0.055
Observations, *n*	1,310,606	1,316,538	1,204,030	96,470	98,676	92,316
Hispanic						
Took course virtually	0.104*** (0.009)	−0.004 (0.006)	−0.035*** (0.007)	−0.026* (0.015)	0.026*** (0.007)	0.043*** (0.007)
Mean of outcome	0.71	0.66	0.79	0.61	0.44	0.69
Proportion taking virtually	0.005	0.005	0.005	0.061	0.061	0.061
Observations, *n*	1,677,931	1,685,618	1,508,533	99,444	101,325	93,604
Asian						
Took course virtually	0.037*** (0.010)	0.008 (0.013)	0.001 (0.011)	0.123*** (0.033)	0.028 (0.035)	0.037 (0.031)
Mean of outcome	0.86	0.79	0.85	0.57	0.43	0.69
Proportion taking virtually	0.011	0.011	0.011	0.120	0.120	0.012
Observations, *n*	137,893	138,079	125,844	3,336	3,371	3,219
Free or reduced-price lunch						
Took course virtually	0.122*** (0.006)	0.001 (0.005)	−0.028*** (0.005)	−0.020* (0.012)	0.011** (0.005)	0.065*** (0.006)
Mean of outcome	0.65	0.60	0.74	0.58	0.41	0.64
Proportion taking virtually	0.006	0.006	0.006	0.061	0.061	0.061
Observations, *n*	2,841,021	2,853,203	2,546,844	192,534	197,028	180,185
Non–free or reduced-price lunch						
Took course virtually	0.136*** (0.005)	−0.012*** (0.004)	−0.027*** (0.004)	0.101*** (0.011)	0.021*** (0.006)	0.063*** (0.005)
Mean of outcome	0.77	0.71	0.81	0.58	0.42	0.67
Proportion taking virtually	0.010	0.010	0.010	0.114	0.114	0.114
Observations, *n*	3,100,403	3,108,526	2,882,249	130,012	132,006	126,072
FCAT Quartile 1						
Took course virtually	0.248*** (0.009)	0.015** (0.007)	−0.008 (0.009)	0.026* (0.014)	0.001 (0.008)	0.084*** (0.008)
Mean of outcome	0.53	0.50	0.70	0.54	0.37	0.62
Proportion taking virtually	0.006	0.006	0.006	0.054	0.054	0.054
Observations, *n*	1,086,471	1,092,671	971,598	101,703	104,414	94,397
FCAT Quartile 2						
Took course virtually	0.178*** (0.008)	−0.003 (0.006)	−0.020*** (0.007)	0.023* (0.013)	0.018** (0.007)	0.050*** (0.006)
Mean of outcome	0.67	0.62	0.79	0.60	0.43	0.68
Proportion taking virtually	0.008	0.008	0.008	0.088	0.088	0.088
Observations, *n*	1,282,967	1,286,643	1,153,017	82,568	84,096	77,559
FCAT Quartile 3						
Took course virtually	0.118*** (0.006)	−0.021*** (0.005)	−0.035*** (0.006)	0.042*** (0.012)	0.004 (0.008)	0.052*** (0.008)
Mean of outcome	0.78	0.72	0.84	0.64	0.47	0.70
Proportion taking virtually	0.010	0.010	0.010	0.120	0.120	0.120

(continued)

TABLE 6 (CONTINUED)

	First-time course takers			Credit recovery		
	1: Pass concurrent course	2: Take/pass follow-on course	3: Graduation proxy	4: Pass concurrent course	5: Take/pass follow-on course	6: Graduation proxy
Observations, n	1,272,471	1,274,738	1,152,648	51,905	52,695	103,811
			FCAT Quartile 4			
Took course virtually	0.054*** (0.004)	−0.039*** (0.005)	−0.043*** (0.005)	0.034** (0.015)	0.044*** (0.011)	0.066*** (0.010)
Mean of outcome	0.90	0.83	0.87	0.70	0.52	0.73
Proportion taking virtually	0.011	0.011	0.011	0.178	0.178	0.180
Observations, n	1,078,015	1,078,920	979,250	17,753	17,968	17,233

Note. Tabulations are based on Florida Department of Education data. Standard errors (in parentheses) are clustered at the school level. Regressions are run separately for each student subgroup. All models include full controls from the final models of Tables 3 and 4. FCAT = Florida Comprehensive Assessment Test.

*$p < .1$. **$p < .05$. ***$p < .01$.

(non–FRPL using, higher achieving) have particularly positive outcomes for our contemporaneous and follow-on course measures. While each of these groups is underrepresented in the course remediation group relative to its numbers in the general first-time course-taker population (Table 2), our results suggest that they benefit more from retaking virtual courses than do their less advantaged peers.

In results presented in the online Appendix D, we show that results are quite consistent for students in different school types as well.

Heterogeneity of Results by Course Subject

We next explore whether the patterns of results that we see are driven by a particular subset of courses or whether they hold across course types. To examine course differences, we separated results by subject: math, science, and language arts (world history, the only social science course in our sample, is presented separately; Table 7).

For first-time course taking, the broad pattern of results for concurrent course taking is similar for all subjects. Estimates of the increased likelihood of course passing associated with virtual course enrollment range from 10.7 percentage points for world history courses to 16.4 percentage points for English language arts courses. Although the contemporaneous coefficient for world history courses is the smallest, it is the only subject area for which virtual course taking is associated with a positive change in the likelihood of taking and passing a follow-on course (column 2, $b = 0.016$). In contrast, students are 1.6 to 2.5 percentage points less likely to take and pass follow-on courses after taking first attempts virtually in English and science ($p < .01$ for both). Estimates for our proxy of likely graduation are negative across all subject areas ($p < .01$).

As with the results for initial attempts, the results for students retaking courses show larger increases in the

likelihood of contemporaneous course passing in English ($b = 0.062$, $p < .001$) and science ($b = 0.107$, $p < .001$) than in math and world history, where coefficients were small in magnitude and nonsignificant. Retaking courses in the virtual setting is not associated with significant changes in the likelihood of jointly taking and passing a future course in math or English. The positive association between virtual credit-recovery attempts and downstream course success is evident only for science and world history courses. Estimates for our graduation proxy measure are positive across all subject areas ($p < .01$) for virtual credit recovery classes. Taken together, these results suggest that virtual science courses offer somewhat more consistent advantages to course retakers than do virtual courses in other subjects, although the positive graduation results hold across all subjects.

Discussion

Online course taking is expanding rapidly for high school students. Florida requires all high school students to take at least one course virtually before graduation, and other states have or are considering similar policies. The online setting offers potential benefits. For students in schools with limited course offerings, the online setting can expand access to curricula. For students seeking additional course taking in the summer—to retake classes in which they performed poorly or to free up time during the school year for other activities—the online setting could again increase opportunities. Yet, prior evidence in K–12 and in higher education comparing students in virtual courses and those in in-person classrooms tended to find that the online setting is currently less effective. Students, especially those with lower prior performance, tend to learn less in the online setting.

In this study, we use data from the state of Florida to compare students in online and in-person classes. In comparison with the prior literature, our study examined an unusually

TABLE 7
Heterogeneity by Course Type

	First-time course takers			Credit recovery		
	1: Pass concurrent course	2: Take/pass follow-on course	3: Graduation proxy	4: Pass concurrent course	5: Take/pass follow-on course	6: Graduation proxy
	Mathematics courses					
Took course virtually	0.109*** (0.007)	−0.001 (0.005)	−0.014*** (0.004)	0.007 (0.013)	−0.0003 (0.006)	0.048*** (0.005)
Mean of outcome	0.66	0.60	0.79	0.54	0.39	0.69
Proportion taking virtually	0.009	0.009	0.009	0.067	0.067	0.067
Observations, *n*	1,651,965	1,658,154	1,482,884	139,981	142,691	130,974
	Science courses					
Took course virtually	0.136*** (0.007)	−0.025*** (0.005)	−0.037*** (0.004)	0.107*** (0.013)	0.047*** (0.008)	0.079*** (0.007)
Mean of outcome	0.74	0.68	0.80	0.57	0.42	0.66
Proportion taking virtually	0.008	0.008	0.008	0.093	0.093	0.093
Observations, *n*	1,857,689	1,863,720	1,674,398	71,462	72,840	68,730
	English courses					
Took course virtually	0.164*** (0.006)	−0.016*** (0.005)	−0.050*** (0.005)	0.062*** (0.011)	−0.006 (0.007)	0.087*** (0.007)
Mean of outcome	0.69	0.63	0.73	0.63	0.43	0.57
Proportion taking virtually	0.008	0.008	0.008	0.088	0.088	0.088
Observations, *n*	1,362,286	1,367,309	1,254,768	72,320	73,901	68,875
	World history courses					
Took course virtually	0.107*** (0.007)	0.016** (0.008)	−0.025*** (0.006)	−0.021 (0.014)	0.032*** (0.010)	0.048*** (0.008)
Mean of outcome	0.77	0.72	0.80	0.63	0.45	0.64
Proportion taking virtually	0.009	0.009	0.009	0.109	0.109	0.110
Observations, *n*	1,069,484	1,072,546	1,017,043	38,783	39,602	37,678

Note. Tabulations are based on Florida Department of Education data. Standard errors (in parentheses) are clustered at the school level. Regressions are run separately for each subject subgroup. All models include full controls from the final models of Tables 3 and 4.
**$p < .05$. ***$p < .01$.

large and diverse set of students and online courses. Moreover, we were able to examine not only current course performance but also future course taking and performance in those courses, as well as proxies for graduation eligibility by Grade 12. To the extent that online course taking increases access to courses that help students progress through school, the effects of online courses could be positive even if learning opportunities are not as great.

Using a rich set of controls and a variety of fixed effects to reduce potential biases, we find that students tend to receive higher grades in online courses. These better grades could be due to better student performance or to more transitory factors, such as easier grading standards. To focus on the effects on performance, we also look at longer-term outcomes. Students taking courses for the first time tend to see less positive longer-term effects when taking courses virtually. When compared with their same-school peers, virtual students are less likely to persist in school through the final

term of a projected Grade 12 year and are marginally less likely to take and pass the next same-subject course in the high school sequence. However, students retaking courses that they had previously failed see some benefits from online course taking, being more likely to pass the contemporaneous course, more likely to take and pass the follow-up course, and more likely to persist through 12th grade.

The differences in estimated effects between first-time takers and retakers could be due to differences in the counterfactual course offerings. First-time takers may have access to a similar course in their school at the time when they are taking it and with students in their cohort, while retakers might be limited in alternatives if they want to retake a given course and maintain their progress through high school—for example, by taking the course in the summer or during nonschool hours. In our data, about 15% of virtual retake efforts come during summer terms; first-time virtual course attempts during the summer are negligible. The access offered by online

courses may enable retakers to progress in ways that are more difficult in brick-and-mortar schools.

Differences between first-time takers and retakers in the effects of online course taking could also be driven by differences in the characteristics of students taking the two types of courses if the effects of online courses are heterogeneous. In part to address this possibility, we estimated effects separately for groups of students, courses, and schools. Overall, the results were largely though not completely consistent across groups. For first-time course takers, virtual course effects were quite consistent across groups, with some evidence of more negative long-run effects for more affluent, higher-scoring, and White students. Regarding course retaking, the effects again are similar across groups, with evidence of more benefits for nonpoor students and students with higher prior achievement. Given that first-time course takers are not more likely to be poor or low achieving than the course retakers, the more positive results for course retakers relative to first-time takers is unlikely to be due to differences in the students served.

Our results differ from those of some past studies in important ways. At first blush, our credit recovery results seem considerably more positive than the results of online credit recovery in previous randomized controlled trials (Heppen et al., 2017). However, our results for math specifically—the set of results most closely related to those of Heppen et al. (2017)—show null effects of online remediation in math on the contemporaneous course-passing outcome and on the joint likelihood of taking and passing a same-subject follow-on course. Coupled with the fact that Heppen and colleagues' trial focused on a lower-income population—a population for which we find modestly negative contemporaneous course passing results in subgroup analyses—the differences in our results seem likely to be driven by the differences in the breadth of subjects and the differences in populations studied. The focus on this broader set of subjects and students is one key contribution of our study.

Our estimates have several limitations that provide the impetus for additional research. We attempt to address selection bias in multiple ways, including by applying a highly saturated set of controls and confirming the robustness of our results to multiple econometric techniques to address selection. Moreover, we calculate that selection on unobservables would have to be substantially greater than on observables to render our results null: for our most conservative estimates, the selection on unobservables would have to exceed selection on observables by 2.5 times to nullify results. However, we know little about the generally-unobserved factors that motivate online enrollment in high school, particularly in the core subject areas in our sample. Additional research could uncover factors that prompt students to take online enrollments in core subjects to help contextualize whether characteristics unobserved in our data may explain away our estimates. More generally, our results

should be interpreted with some caution given that the number of students taking online courses in these core subjects remains relatively small.

Finally, our estimates are based on a specific point in time—2006–2007 through 2011–2012—yet the effects of online course taking are unlikely to be static. As students develop greater comfort with the online setting, their performance could easily change. The nature of virtual courses may change over time as well: Artificial intelligence, which is just beginning to penetrate online courses, offers promise for improving online learning by responding to students' abilities and personalizing instruction in ways that classroom teachers would struggle to replicate and by learning which types of information and assessments best facilitate student learning. To support our analytic strategy, our sample focused on courses commonly available in home institutions, but this required us to leave unexamined the potential for online courses to expand access to new courses that are less commonly offered; future researchers should look into this possibility to further evaluate claims that online courses offer the possibility that geography will cease to determine access to quality teaching and diverse course offerings. Nonetheless, we provide an important first set of evidence that these benefits are yet to be fully realized.

Notes

1. While the effects of online courses relative to face-to-face courses has tended to be negative or neutral, especially for students with poorer prior performance, these effects are for students who would take the courses in either format, depending on which one was available. Online courses may open up course taking to students who would not take in-person courses, such as those who live far away from brick-and-mortar institutions or those with unpredictable or variable work hours whose schedules do not fit with traditional course requirements. We do not weigh these potential benefits in this article.

2. The full-time FLVS programs are open to all students. The part-time program is open to all students who attended a Florida public school in the prior academic year, to K–1 students, to students whose siblings are currently enrolled in virtual schools or who were enrolled at the end of the last academic year, to military dependents who recently moved to Florida, and to K–5 students currently enrolled in full-time virtual programs.

3. Including the part-time school, the full-time school, and district franchises.

4. This may happen, for instance, if students drop out without retaking the course, if they take the course more than a year after the initial failure, or if they take a different course to fill graduation requirements. In the online Appendix A, we address these possibilities and discuss how they affect our sample.

5. A small share of students takes multiple courses in the same subject simultaneously (e.g., a physics class and a chemistry class following biology). In those cases, we prioritize the grade from the course that is most often taken following the original course. Where there is no clear course to prioritize (in 6% of cases), we randomly select one of the follow-on courses; excluding cases where we made these random determinations has little impact on results.

6. An additional set of outcomes of interest would be test scores. Unfortunately, the match rates for end-of-course tests are too low to enable us to estimate effects reasonably. Our match rates for ninth- and 10th-grade math and reading FCAT tests are better (although FCAT availability changes over the years covered in our sample). Like Chingos and Schwerdt (2014), we find modest advantages for virtual students over face-to-face students on these outcomes, although our match rates for FCAT scores are lower for virtual students than for face-to-face students, making these results difficult to interpret cleanly. Results are available on request.

7. This requirement could be fulfilled in sixth to eighth grade as well as in high school. It could also be fulfilled with enrollments at K–12 online schools such as FLVS or through dual enrollment taken through a postsecondary institution (Florida Statute 1002.4282, 2018).

8. Specifically, 90% of randomized results and 45% of nonrandomized results survive this test (Oster 2017).

9. The results are similar if we do not limit the adjustment factor to apply only to student selection.

10. Our R^2 values are lower in general than Oster's (2017): the highest R_{max} in the models that we present is around 0.27, while her R_{max} values are generally >0.5. This might be partially due to differences in the types of outcome variables estimated; her examples use continuous outcomes, while our outcomes are dichotomized. To look into this further, we reestimate these models using contemporaneous course grades (on a 1–4 scale, standardized), which is the only one of our outcomes that lends itself easily to reestimation with a continuous outcome. For first-time course takers and credit recovery students, δ values remain well above the 1.0 threshold (δ = 27.870 for first-time course takers, δ = 8.186 for credit recovery). Results are available on request. We thank an anonymous reviewer for raising this point.

ORCID iD

Michael Hill ⓘD https://orcid.org/0000-0001-7699-7095

References

Ahn, J., & McEachin, A. (2017). Student enrollment patterns and achievement in Ohio's online charter schools. *Educational Researcher*, *46*(1), 44–57.

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, *113*(1), 151–184.

Barbour, M. K. (2013). The landscape of K–12 online learning: Examining what is known. In M. G. Moore (Ed.), *Handbook of distance education* (3rd ed., pp. 574–593). New York, NY: Routledge.

Berge, Z. L., & Clark, T. (2005). *Virtual schools: Planning for success*. New York, NY: Teachers College Press.

Bettinger, E., Fox, L., Loeb, S., & Taylor, E. (2017). Virtual classrooms: How online college courses affect student success. *American Economic Review*, *107*(9), 2855–2875.

Bork, R., & Rucks-Ahidiana, Z. (2013, October). *Role ambiguity in online courses: An analysis of student and instructor expectations* (Working Paper No. 64). Retrieved from the Community College Research Center website: http://ccrc.tc.columbia.edu /media/k2/attachments/role-ambiguity-in-online-courses.pdf

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

Carrell, S. E., & West, J. E. (2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy*, *118*(3), 409–432.

Cavalluzzo, L., Lowther, D. L., Mokher, C., & Fan, X. (2012). *Effects of the Kentucky Virtual Schools' hybrid program for Algebra I on Grade 9 student math achievement*. Washington, DC: Institute of Education Sciences.

Chingos, M. M., & Schwerdt, G. (2014, September). Virtual school and student learning: Evidence from Florida Virtual Schools. *Retrieved from Program on Education Policy and Governance website:* http://www.hks.harvard.edu/pepg/PDF/FLVS%20 PEPG%20working%20paper%20%283%29.pdf

Evergreen Education Group. (2015). *Keeping pace with K–12 digital learning: An annual review of policy and practice* (12th ed.). Durango, CO: Author.

Figlio, D. N., Rush, M., & Yin, L. (2013). Is it live or is it internet? Experimental estimates of the effects of online instruction on student learning. *Journal of Labor Economics*, *31*(4), 763–784.

Figlio, D. N., Schapiro, M. O., & Soter, K. (2015). Are tenure track professors better teachers? *Review of Economics and Statistics*, *97*(4), 715–724.

Florida Department of Education. (2016a). *Secondary student progression: 2016–2017 frequently asked questions*. Retrieved from http://www.fldoe.org/core/fileparse.php/7764/urlt/Secondary _Student_Progression_FAQs.pdf

Florida Statute 1002.321. (2018). *Digital learning*. Retrieved from http://www.leg.state.fl.us/statutes/index.cfm?App_mode =Display_Statute&Search_String=&URL=1000–1099/1002 /Sections/1002.321.html

Florida Statute 1002.4282. (2018). *Requirements for a standard high school diploma*. Retrieved from http://www.leg.state.fl .us/statutes/index.cfm?App_mode=Display_Statute&Search_ String=&URL=1000–1099/1003/Sections/1003.4282.html

Florida Virtual School. (n.d.). *Student/teacher ratio*. Retrieved from https://www.flvs.net/about/instruction/student-teacher -ratio

Friend, B., & Johnson, S. (2005). Florida Virtual School: A choice for all students. In Z. L. Berge (Ed.), *Virtual schools: Planning for success* (pp. 97–117). New York, NY: Teachers College Press.

Gonzalez, S. (2012, August 14). No summer school means some students repeat a grade this fall. *StateImpact*. Retrieved from https://stateimpact.npr.org/florida/2012/08/14/no-summer -school-means-some-students-repeat-a-grade-this-fall/

Hart, C. M., Friedmann, E. A., & Hill, M. (2018). Online course-taking and student outcomes in California community colleges. *Education Finance and Policy*, *13*(1), 42–71.

Heissel, J. (2016). The relative benefits of live versus online delivery: Evidence from virtual Algebra I in North Carolina. *Economics of Education Review*, *53*, 99–115.

Heppen, J., Sorensen, N., Allensworth, E., Walters, K., Rickles, J., Stachel Taylor, S., & Michelman, V. (2017). The struggle to pass algebra: Online vs. face-to-face credit recovery for at-risk urban students. *Journal of Research on Educational Effectiveness*, *10*(2), 272–296.

Jacob, B., Berger, D., Hart, C., & Loeb, S. (2016). Can technology help promote equality of educational opportunities? *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *2*(5), 242–271.

Joyce, T. T., Crockett, S., Jaeger, D. A., Altindag, O., & O'Connell, S. D. (2015). Does classroom time matter? *Economics of Education Review*, *46*, 64–77.

Krieg, J. M., & Henson, S. E. (2016). The educational impact of online learning: How do university students perform in subsequent courses? *Education Finance and Policy*, *11*(4), 426–448.

Means, B., Bakia, M., & Murphy, R. (2014). *Learning online: What research tells us about whether, when, and how*. New York, NY: Routledge.

Means, B., Toyama, Y., Murphy, R., & Bakia, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record*, *115*(3), 1–47.

National Center for Education Statistics. (2012). Table 121. *Digest of Education Statistics*. Retrieved from https://nces.ed.gov /programs/digest/d12/tables/dt12_121.asp

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*. doi:10.1080/07350015.2016.1227711

Sheehy, K. (2012, October 24). *States, districts require online ed for high school graduation*. Retrieved from https://www .usnews.com/education/blogs/high-school-notes/2012/10/24 /states-districts-require-online-ed-for-high-school-graduation

Streich, F. (2014). Online education in community colleges: Access, school success, and labor market outcomes (Doctoral dissertation). University of Michigan, Ann Arbor.

Tallent-Runnels, M. K., Thomas, J. A., Lan, W. Y., Cooper, S., Ahern, T. C., Shaw, S. M., & Liu, X. (2006). Teaching courses online: A review of the research. *Review of Educational Research*, *76*(1), 93–135.

Woodworth, J. L., Raymond, M. E., Chirbas, K., Gonzalez, M., Negassi, Y., Snow, W., & Van Donge, C. (2015). *Online charter school study*. Stanford, CA: Center for Research on Education Outcomes.

Xu, D., & Jaggers, S. S. (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), 360–377.

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

Authors

CASSANDRA M. D. HART is an associate professor in the School of Education at the University of California, Davis. Her research interests include online education, school choice, and teacher-student race match effects.

DAN BERGER is a doctoral candidate at the University of Michigan. His research interests include program evaluation and school data use.

BRIAN JACOB is the Walter H. Annenberg Professor of Education Policy, Professor of Economics, and Professor of Education at the University of Michigan. His current research focuses on school accountability, teacher labor markets, and virtual schooling.

SUSANNA LOEB is the director of the Annenberg Institute and professor of education and international and public affairs at Brown University.

MICHAEL HILL is a doctoral candidate in the School of Education at the University of California, Davis. His research focuses on student outcomes and instructional strategies in online and technology-mediated educational environments.