Educational Evaluation and Policy Analysis
September 2011, Vol. 33, No. 3, pp. 360–377
DOI: 10.3102/0162373711413814

The Effectiveness of Distance Education Across Virginia's Community Colleges: Evidence From Introductory College-Level Math and English Courses

Di Xu
Shanna Smith Jaggars
Columbia University

Although online learning is rapidly expanding in the community college setting, there is little evidence regarding its effectiveness among community college students. In the current study, the authors used a statewide administrative data set to estimate the effects of taking one's first college-level math or English course online rather than face to face, in terms of both course retention and course performance. Several empirical strategies were used to minimize the effects of student self-selection, including multilevel propensity score. The findings indicate a robust negative impact of online course taking for both subjects. Furthermore, by comparing the results of two matching methods, the authors conclude that within-school matching on the basis of a multilevel model addresses concerns regarding selection issues more effectively than does traditional propensity score matching across schools.

Keywords: *distance education, community college, propensity score, multilevel design*

In the past decade, distance education through online coursework[1] has become a common option for students in higher education: Over 29% of U.S. college students took online courses in the fall of 2009, and the 19% annual growth rate in online enrollment over the past decade far exceeds the growth rate in overall higher education enrollment (Allen & Seaman, 2010). Advocates of distance education have noted several potential benefits of online learning in comparison to the traditional face-to-face format. Online courses offer the flexibility of off-site asynchronous education (Peterson & Bond, 2004) and have the potential to provide strong computer-mediated student-to-student interaction and collaboration (Cavus & Ibrahim, 2007; Harasim, 1987), as well as immediate feedback on student learning (Brown, Lovett, Bajzek, & Burnette, 2006).

Advocates are also particularly optimistic about the potential of fully online coursework to improve and expand learning opportunities at community colleges, which educate large proportions of nontraditional students (Choy, 2002; Kleinman & Entin, 2002). These students, who may find it difficult to attend on-campus courses because of employment or family commitments, are more likely than traditional students to choose online learning (Cohen & Brawer, 2003; Imel,

We would like to thank all the staff at the Community College Research Center for their support during this research project. We are also indebted to Jennifer Hill, Thomas Bailey, Judith Scott-Clayton, Nikki Edgecombe, and Sung-Woo Cho for their valuable comments and suggestions on this article. The 23 community colleges in Virginia provided invaluable data on which this article is based. In addition, we are grateful to the Lumina Foundation and the Bill & Melinda Gates Foundation, which have provided crucial funding for this research. All opinions and mistakes are our own.

1998; Perez & Foshay, 2002). Accordingly, online learning enrollments have increased more quickly at 2-year colleges than at 4-year colleges in the past decade (Choy, 2002; Parsad & Lewis, 2008). By 2007, over 97% of 2-year colleges offered online courses, compared with 66% of all postsecondary institutions (Parsad & Lewis, 2008).

Despite the rapid growth of and high hopes for distance education in community colleges, questions remain regarding its effectiveness in this particular educational setting. Although the "no significant difference" phenomenon between face-to-face and distance education described by Russell (2001) continues to dominate the literature, most studies in this area focus on university students who are academically well prepared (Jaggars & Xu, 2010). As a result, there is little evidence on the effectiveness of online courses among community college students, the majority of whom enter college academically underprepared (Attewell, Lavin, Domina, & Levey, 2006; Bailey, Jeong, & Cho, 2010). Some evidence, however, suggests that students who are less prepared may struggle with online coursework. A recent experimental study comparing learning outcomes between online and face-to-face sections of an economics course (Figlio, Rush, & Yin, 2010) found no significant difference between the two groups overall but noted that among students with low prior grade point averages (GPAs), those in the online condition scored significantly lower on in-class exams than did those in the face-to-face sections. Observational studies (e.g., Bendickson, 2004; Chambers, 2002; Vargo, 2002) also suggest that online courses are associated with less desirable course outcomes for underprepared students.

In addition, the bulk of research comparing online and face-to-face courses has focused on learning outcomes among those who complete the courses, paying little attention to course completion. Yet online courses are often observed to have higher midsemester withdrawal rates than face-to-face courses (e.g., Beatty-Guenter, 2003; Carr, 2000; Chambers, 2002; Cox, 2006; Moore, Bartkovich, Fetzner, & Ison, 2003). This difference may be particularly pronounced among underprepared students. For example, one study of community college students in developmental mathematics observed that 73% of face-to-face students completed the course with a grade of A, B, or C, whereas only 51% of online students did so (Summerlin, 2003). These observed trends are worrisome, given that course completion is a fundamental measure of success for community college students. Students who withdraw from a course midsemester run the very real risk of never returning to successfully complete the course, thereby prohibiting progression to the next course in the sequence (see, e.g., Bailey et al., 2010). Moreover, many community college students have low incomes (Adelman, 2005) and can ill afford to pay full tuition for courses that they do not successfully complete. However, some practitioners and researchers argue that high online withdrawal rates are due not to the course format but to "self-selection bias"; that is, the type of student who chooses to enroll in an online course is also the type of student who is more likely to withdraw (Howell, Laws, & Lindsay, 2004; Hyllegard, Heping, & Hunter, 2008). Thus, it is important to examine whether, after controlling for student characteristics, online courses continue to have higher attrition rates than face-to-face courses.

Moreover, it is critical to take into account the potential of differential attrition when examining course learning outcomes. For example, in a study of developmental writing students at a midwestern community college, Carpenter, Brown, and Hickman (2004) controlled for a variety of factors and found that students were significantly more likely to withdraw from an online course than from a face-to-face course. However, online students who completed the course earned higher grades than face-to-face completers, net of controls. The authors acknowledged that the grade effect could be due to the substantially higher withdrawal rates in the online sections. When examining withdrawal patterns more closely, they found that students with lower placement scores were more likely to withdraw from the online section, while students with higher scores were more likely to withdraw from the face-to-face section, leaving the online section with better prepared students. This pattern gives weight to the notion that differential withdrawal rates can result in misleading comparisons between students who complete online and face-to-face courses. Thus, most existing studies not only ignore student retention as an outcome of interest but in doing so may also introduce bias into their examinations of course performance, because students may withdraw from different course formats at different rates and for different reasons.

In the current study, we examine the effectiveness of taking one's first college-level math and English courses online (rather than face to face) within the Virginia Community College System (VCCS). Introductory college-level courses in math and English represent essential prerequisites for most degrees and certificates and as such are commonly termed "gatekeeper" courses. We focus on these gatekeeper courses for both applied and methodological reasons. From an applied standpoint, the successful completion of gatekeeper courses plays a critical role in one's college career; passing these initial college-level courses results in a substantially higher probability of earning a postsecondary credential (Calcagno, Crosta, Bailey, & Jenkins, 2007). As a result, community colleges tend to be particularly concerned with success rates in these courses and how to improve them. From a methodological standpoint, these courses have very high enrollments compared with more advanced college courses, yielding a large sample size for analysis. In addition, most students take these courses in early stages of their college careers, when they are less likely to have preexisting knowledge regarding online courses in their colleges or their likely performance in these courses. Accordingly, focusing on these introductory courses (rather than more advanced courses) should reduce self-selection bias.

To assess the effects of taking a gatekeeper course online, we explored two course outcomes: (a) midsemester course withdrawal (also termed *attrition* or *dropout*) and (b) the likelihood of earning a grade of C[2] or better. To address potential self-selection bias, we used three different regression techniques, including multilevel propensity score matching to account for school-level variation in online enrollment (discussed further in the "Empirical Framework and Methodology" section).

The current study makes four contributions to the existing literature on distance learning in higher education. First, we focus on the understudied sector of community colleges. Second, we explicitly examine course completion and adjust for differential completion in our examination of learning outcomes. Third, most previous studies have explored the "average treatment effect" of online course taking, comparing online students with a heterogeneous group of face-to-face students (some of whom would be likely to take online courses and others of whom would be very unlikely to do so). Yet policymakers are more concerned about the effects of online learning on the type of student who is likely to take online coursework, which might be substantially different from the average effect on the overall student population. In this study, we use propensity matching methods to estimate the effect of treatment on this subset of students in addition to the larger student population. Fourth, using a statewide data set of individuals attending multiple institutions, we estimate the impact of online learning across a large higher education system. Although focusing on only one state might limit the external validity of our results, Virginia's system structure, population demographics, and application of online courses are similar to those seen nationally, so our results may be reasonably considered to reflect on distance education issues facing many other community college systems.

Our analyses, discussed in more detail below, show robust estimates of a negative impact of online learning on course retention and course performance in both math and English. Propensity score estimation indicated that online course takers are substantially different from classroom-based students in terms of an array of characteristics; in particular, online takers tend to have stronger academic preparation. Thus, simple comparisons between online and face-to-face students may underestimate the negative impacts of the online format on course outcomes. Indeed, in this study, the raw observed differences translate into much larger effect sizes after controlling for baseline characteristics.

The remainder of this article is organized as follows: First, we describe the VCCS database. We then introduce our empirical strategies, with a focus on the multilevel propensity score matching method. Next we present the results of propensity score analyses, examining course persistence and course performance, and we then present the results of robustness checks and sensitivity analyses. Finally, we summarize the results and present recommendations.

Data Description

VCCS Data and Institutional Characteristics

Analyses were performed on a data set containing nearly 24,000 students from 23 community colleges in Virginia. First-time students who initially enrolled during the summer or fall of 2004 were tracked until the summer of 2008, for approximately 4 years. The data set contained information on student demographics, institutions attended, transcript data on courses taken and grades received, and information on educational attainment. Information on each course was also included, such as the course subject, whether it was a developmental or college-level course, and whether it was a distance education or face-to-face section.[3] Students who dropped the course early in the semester (prior to the course census date) are not included in the data set. Thus, in our data set, a dropout student paid full tuition for the course but did not persist to the end of the course.

The 23 Virginia community colleges vary widely from one another in terms of institutional characteristics. The system comprises a mix of large and small schools, as well as institutions located in rural, suburban, and urban settings. For example, the system contains a large multicampus institution with a high proportion of minority students located in the suburbs of a major metropolitan area, but it also contains several small, rural, predominantly White schools. Overall, however, Virginia community colleges seem to represent a rural, low-income, underfunded, and African American student population.[4]

Gatekeeper Courses in the VCCS

In terms of English, all 23 schools share the same introductory composition course (ENG111), which serves as a gatekeeper for all students enrolled in 1- or 2-year programs of study. In terms of math, the required first college-level course varies by program; the list of gatekeepers includes algebra, precalculus, calculus, quantitative reasoning courses, and discipline-specific courses such as business mathematics. Across the sample, 61% of students (n = 13,973) enrolled in English gatekeeper courses and 37% (n = 8,330) enrolled in math gatekeeper courses at some point in their VCCS careers (“gatekeeper takers”). Twenty-three percent of the English gatekeeper takers and 38% of the math gatekeeper takers first took corresponding remedial courses because of a lack of academic preparation.[5] Although the majority of students took only one gatekeeper course in a given subject area, 14% of the English takers attempted ENG111 more than once, and 33% of the math takers reattempted the same course or later took an additional gatekeeper course. For these students, only the first gatekeeper course attempt in the given subject area was used for the current analysis. Most students took their first gatekeeper courses within the first year (84% for English and 62% for math); however, some took them in later terms. As a result, time dummies were included in both the propensity score estimation model and the analytical model to account for possible time variations in treatment, outcome, and the effects of treatment on outcome.

Consistent with the wide variation in enrollment size across the 23 community colleges, the number of students taking gatekeeper courses also varied across schools. Across colleges, the English gatekeeper sample ranged from 93 to 3,515 and the math gatekeeper sample from 42 to 2,310. Colleges also varied in terms of students’ course persistence (ranging from 74% to 96% for English and from 79% to 97% for math) and, among students who completed the courses, in terms of performance (ranging from 67% to 86% for English and from 59% to 83% for math). The substantial school-level variation in course outcomes may depend in part on observed school characteristics, such as geographic location, school size, or institutional resources (see the following section for details), and in part on unobserved characteristics, such as instructional quality, class size, and school managerial effectiveness. These patterns highlight the necessity of controlling for school-level variation in the current study.

Online Courses in the VCCS

Each college in the VCCS has developed its online program locally, according to the institution’s own priorities, resources, and the perceived needs of its particular student population (as do most community colleges; see Cox, 2006). Accordingly, colleges varied widely in the proportion of courses that their students took online. In terms of ENG 111, the percentage of students who took the course online varied across colleges from a minimum of 1% to a maximum of 55%. Similarly, in gatekeeper math, the proportion of enrollments taken online varied across colleges from 1% to 43%.

In general, across the 4-year period of the study, online course taking increased dramatically in Virginia's community colleges. In the fall of 2004, entering students attempted an average of only 0.57 credits (6% of their semester credits) online; by the spring of 2008, still-enrolled students in the 2004 cohort had tripled the rate of distance credit attempts, to an average of 1.72 credits (21% of their semester credits). As we discuss in a separate report (Jaggars & Xu, 2010), this growth was due to two separate trends. First, students in the VCCS 2004 cohort were increasingly likely to try at least one online course across time. Second, when considering only students who were actively online in a given semester, the percentage of credits taken online also increased across semesters.

Empirical Framework and Methodology

Logistic Regression Model Estimation

To assess how course outcomes differ between online takers and similar face-to-face takers, we used techniques based on logistic regression. To control for school-level variance in course offering as well as course outcomes, we used multilevel logistic regression.[6] Taking course attrition as an example outcome measure (equation 1),

$$\begin{aligned} \text{logit}(D_{ij}) &= \beta_{0j} + \beta_{1j}\mathbf{X}_{ij} + \mu_{ij} \\ \beta_{0\text{j}} &= \gamma_{00} + \gamma_{01}\mathbf{W}_j + \varepsilon_{0j};\ \beta_{1j} = \gamma_{10}. \end{aligned} \tag{1}$$

In this set of equations, D_{ij} is the course dropout outcome for student i in school j and is equal to 1 if the student dropped the course; $\mathbf{X}_{ij}$ is a vector of individual-associated baseline variables, which include demographic characteristics, academic performance indicators, and the timing when the student took the course (Table 1); $\mathbf{W}_j$ is a vector of school-level variables; μ_{ij} is the individual-level error term that captures unobserved variation between students within schools; and ε_{0j} is the school-level error term that captures unobserved variation across schools. The multilevel logistic regression model explores the average treatment effect of online course taking, comparing online students with a heterogeneous group of face-to-face students, some of whom are very unlikely to take an online course. To estimate the impact of online learning on the type of student who is likely to take online coursework, we used propensity score matching.

Propensity Matching Estimation

Matching across schools versus matching within schools. Propensity score matching is widely used to draw sound inferences in observational studies. In the current study, if we assume that the course selection process is the result of only observable covariates denoted by $\mathbf{X}$, we can use the propensity score (Rosenbaum & Rubin, 1985), defined as the conditional probability of receiving a treatment given pretreatment characteristics $\mathbf{X}$, to simulate a comparison group of face-to-face takers who resemble the online takers of a gatekeeper course. Therefore, as long as there are no unobserved confounders independent of $\mathbf{X}$, comparisons of course outcomes on the basis of the matched sample allow us to estimate the treatment effect of the online format on those who took their first English or math gatekeeper courses online. The propensity score strategy has two major advantages in the current research scenario. Practically, this estimate directly addresses the policy concern of most community colleges, by focusing only on potential online takers. Methodologically, propensity matching can better address the selection issue by making inferences only from data on students who are similar on observed characteristics.

However, a key challenge in applying the propensity score method in multisite investigations is that the average log odds of receiving the treatment may vary across schools (Arpino & Mealli, 2008; Kim, 2006) and that this variation may depend on measured or unmeasured school-level characteristics. To control for school-level variation in online course offering and enrollment, we used multilevel

TABLE 1

Balance Between Online and F2F Course Takers (sample: English gatekeeper takers)

		Unmatched				Matched across schools				Matched within schools			
Variable	Sample	*M*[a]	*s*	*SD*[b]	*s* Ratio[c]	*M*	*s*	*SD*	*s* Ratio	*M*	*s*	*SD*	*s* Ratio
Demographic variables													
Age	Online	23.97	7.92	0.47	1.43	23.95	7.98	0.07	1.03	23.76	7.82	0.03	1.06
	F2F	20.76	5.53			23.41	7.76			23.98	8.38		
Age^2	Online	637.24	488.48	0.42	1.47	636.00	487.06	0.01	1.04	625.59	482.47	0.00	1.10
	F2F	461.63	332.93			639.01	467.60			625.41	529.01		
Female	Online	0.70	0.46	0.25	1.06	0.70	0.50	0.02	1.09	0.70	0.46	0.00	1.00
	F2F	0.58	0.49			0.71	0.46			0.70	0.46		
Black	Online	0.16	0.37	0.08	1.05	0.16	0.37	0.05	1.03	0.17	0.37	0.00	1.03
	F2F	0.19	0.39			0.18	0.38			0.17	0.38		
American Indian	Online	0.01	0.08	0.00	1.14	0.01	0.08	0.00	1.13	0.01	0.08	0.00	1.25
	F2F	0.01	0.07			0.01	0.09			0.01	0.10		
Asian	Online	0.03	0.18	0.18	1.39	0.03	0.18	0.06	1.00	0.03	0.18	0.00	1.00
	F2F	0.07	0.25			0.04	0.18			0.03	0.18		
Hispanic	Online	0.03	0.16	0.15	1.43	0.03	0.16	0.07	1.23	0.03	0.16	0.00	1.13
	F2F	0.06	0.23			0.02	0.13			0.03	0.18		
Age at college entry	Online	0.31	0.46	0.48	1.44	0.31	0.46	0.00	1.00	0.30	0.46	0.00	1.00
	F2F	0.12	0.32			0.31	0.46			0.30	0.46		
Transfer-track student	Online	0.55	0.50	0.25	1.06	0.55	0.50	0.02	1.00	0.56	0.50	0.02	1.00
	F2F	0.67	0.47			0.56	0.50			0.55	0.50		
Dual enrolled	Online	0.10	0.29	0.11	1.16	0.10	0.29	0.00	1.03	0.10	0.30	0.00	1.03
	F2F	0.07	0.25			0.10	0.30			0.10	0.29		
ESL student	Online	0.02	0.13	0.12	1.47	0.02	0.13	0.00	1.15	0.02	0.14	0.00	1.07
	F2F	0.04	0.19			0.02	0.15			0.02	0.15		
Federal financial aid student	Online	0.58	0.49	0.41	1.00	0.58	0.49	0.02	1.02	0.57	0.50	0.02	1.02
	F2F	0.38	0.49			0.59	0.48			0.58	0.49		
Took English remedial courses	Online	0.29	0.46	0.14	1.10	0.29	0.46	0.04	1.02	0.29	0.45	0.00	1.02
	F2F	0.23	0.42			0.27	0.45			0.29	0.46		
Full-time student	Online	0.52	0.50	0.20	1.02	0.52	0.50	0.06	1.00	0.52	0.50	0.02	1.00
	F2F	0.62	0.49			0.49	0.50			0.51	0.50		
Took computer literacy course	Online	0.31	0.46	0.26	1.15	0.31	0.46	0.06	1.09	0.31	0.50	0.00	1.06
	F2F	0.20	0.40			0.28	0.50			0.31	0.47		
Credits attempted in the current term	Online	10.37	3.95	0.29	1.09	10.38	3.95	0.07	1.04	10.43	3.92	0.01	1.01
	F2F	11.45	3.61			10.12	3.79			10.45	3.87		

(continued)

TABLE 1. (continued)

		Unmatched				Matched across schools				Matched within schools			
Variable	Sample	*M*[a]	*s*	*SD*[b]	*s* Ratio[c]	*M*	*s*	*SD*	*s* Ratio	*M*	*s*	*SD*	*s* Ratio
Time taking the first English gatekeeper course													
Year 2005–2006	Online	0.17	0.37	0.18	1.19	0.17	0.37	0.03	1.02	0.16	0.37	0.03	1.05
	F2F	0.11	0.31			0.16	0.36			0.17	0.39		
Year 2006–2007	Online	0.06	0.25	0.14	1.47	0.06	0.24	0.00	1.04	0.06	0.24	0.00	1.04
	F2F	0.03	0.17			0.06	0.23			0.06	0.23		
Year 2007–2008	Online	0.04	0.19	0.19	1.58	0.04	0.19	0.05	1.16	0.04	0.18	0.00	1.06
	F2F	0.01	0.12			0.05	0.22			0.04	0.19		
Summer	Online	0.13	0.34	0.24	1.48	0.13	0.33	0.03	1.06	0.12	0.32	0.00	1.03
	F2F	0.06	0.23			0.14	0.35			0.12	0.31		
School variables													
Instructional expenditure per student	Online	3,238.22	544.30	0.21	1.13	3,238.06	544.55	0.02	1.19	3,235.45	544.23	0.00	1.00
	F2F	3,129.50	482.51			3,249.95	648.39			3,235.45	544.23		
Academic expenditure per student	Online	717.47	228.17	0.34	1.26	717.09	227.95	0.00	1.00	714.88	229.45	0.00	1.00
	F2F	646.93	181.68			717.45	228.19			714.88	229.45		
Student expenditure per student	Online	554.04	178.41	0.34	1.83	553.51	177.66	0.04	1.09	555.00	177.95	0.00	1.00
	F2F	505.14	97.75			546.71	163.52			555.00	177.95		
Institutional expenditure per student	Online	1,120.20	336.06	0.32	1.40	1,119.77	335.91	0.02	1.02	1,120.73	338.66	0.00	1.00
	F2F	1,027.10	239.38			1,113.40	328.92			1,120.73	338.66		
Percentage of federal financial aid student	Online	43.45	18.64	0.61	1.15	43.43	18.64	0.01	1.01	43.33	18.74	0.00	1.00
	F2F	32.82	16.18			43.70	18.78			43.33	18.74		
Located in rural area (urban as base group)	Online	0.43	0.50	0.41	1.16	0.43	0.50	0.02	1.02	0.43	0.49	0.00	1.00
	F2F	0.24	0.43			0.42	0.49			0.43	0.49		
School located in suburban area	Online	0.23	0.42	0.31	1.14	0.23	0.42	0.02	1.02	0.23	0.42	0.00	1.00
	F2F	0.37	0.48			0.24	0.43			0.23	0.42		
Percentage of minority student	Online	24.33	15.36	0.45	1.06	24.35	15.35	0.01	1.00	24.50	15.39	0.00	1.00
	F2F	31.09	14.52			24.50	15.32			24.50	15.39		

Note. ESL = English as a second language; F2F = face to face; *s* = standard deviation.

a. Group mean.

b. Standardized difference in group means, calculated following the formula of Austin (2007):

$$SD = \frac{|\overline{X}_{online} - \overline{X}_{F2F}|}{\sqrt{\frac{S_{online}^2 + S_{F2F}^2}{2}}}.$$

c. The ratio of the standard deviation between the online group and the F2F group, calculated by dividing the group with the higher deviation by the standard deviation of the other group.

logistic regression model for propensity score estimation (equation 2):

$$\text{logit}(T_{ij}) = \beta_{0j} + \beta_{1j}\mathbf{X}_{ij} + \mu_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}\mathbf{W}_j + \varepsilon_{0j};\ \beta_{1j} = \gamma_{10}, \qquad (2)$$

where T_{ij} is the treatment assignment for student i in school j and is equal to 1 if the student took the course online. The remaining terms are equivalent to those discussed previously.

Propensities derived from the multilevel model were used to match online and face-to-face students. To find the best match for a given online student, matching could proceed either across schools (i.e., selecting the face-to-face student with the most similar propensity score, regardless of school membership) or within schools (i.e., selecting the face-to-face student with the most similar propensity score within the same school). If the selection process differs between schools, within-school matching may be more appropriate (Arpino & Mealli, 2008). Given the widely different rate of online course uptake across schools, it seems likely that schools differ in their approach to online course development, marketing, and recruitment. To allow for this possibility, and to assess the extent to which it may affect our estimated results, we used both the across- and within-school matching approaches.

Estimation procedures for the propensity matching approach. Estimation was performed in three steps. We describe these steps below in terms of English courses; the same steps were performed separately for math courses. First, for each student in the VCCS sample who ever took ENG111, we estimated the student's propensity to take the course online using equation 2. Second, we used the estimated propensity scores to find the nearest matching face-to-face student for each online student, using the nearest-neighbor method within caliper 0.1 (Dehejia & Wahba, 2002). That is, students in the online group who had no near match (within 0.1 standard deviations of the propensity score) in the face-to-face group were dropped from analysis. Under the across-school matching method, the nearest matching face-to-face student may have attended any school; under the within-school method, the nearest match was required to attend the same school as the online student. For each type of matching method, we then checked whether we had succeeded in balancing the covariates in each school and for the whole sample. The propensity model specifications in the first step were modified several times to achieve a better balance on each potential confounder within each individual school and overall. Balance on both means and higher order sample moments, such as standard deviations for each confounder, were checked in the propensity score model specification process (Hill, 2008). The final models include 28 individual-level variables[7] and 8 school-level variables[8] (see Table 1).

Given that there were two outcome measures (course persistence and, among those who persisted, course performance), the matching processes were conducted separately for each of the two outcome variables. For example, for the ENG111 analysis, we first matched all online ENG111 students with face-to-face students holding the closest propensity scores and used this matched sample for the analysis on course retention. We then matched online takers who were retained through ENG111 with students retained face to face and used this new matched sample for the analysis of course performance, with successful performance defined as earning a grade of C or better. This separate matching process addresses the concern that students might withdraw from the online and face-to-face sections at different rates and for different reasons, thus leaving two groups of students with different baseline characteristics. Overall, given that we had two types of courses (math and English), two types of matching methods (across and within schools), and two types of outcomes (persistence and performance), we performed eight separate matching processes, resulting in eight distinct matched samples for analysis.

In the third and final step of the analysis, the treatment effect of online learning on the given outcome was estimated separately on each of the eight samples. To reduce bias in postmatching analysis, many researchers (Abadie & Imbens, 2002; Hill, 2008; Rosenbaum & Rubin, 1985; Rubin & Thomas, 2000) recommend performing additional covariance adjustment through stratification or regression modeling. In our application, we estimated the treatment effects on the basis of the matched data using a multilevel logistic regression with random intercepts, where the treatment indicator and all confounders were included in the postmatch analysis to increase the precision of the treatment effect estimators (see equation 1).

Assumptions and their plausibility. To provide an unbiased estimate of the average treatment effect on the "treated" (online takers), three major assumptions are required. First, the inference made through propensity score matching is based on the ignorability assumption. Although we used within-school matching to control for potential confounders at the school level and had access to rich individual information in developing the propensity model, we could not rule out the possibility that the ignorability assumption might be violated. To address this concern, we conducted further sensitivity analysis, the results of which are presented in the "Robustness Check and Sensitivity Analysis" section below. Second, there should be no interference between units: The treatment assignment for one unit should not affect the outcome for another. This is the stable unit treatment value assumption. Given that online students and face-to-face students enroll in different sections of a course, they are unlikely to affect each other, which generally satisfies this assumption. Third, the analysis assumes that matching is performed over the common support region. Given that the online taker population is small compared with face-to-face takers, and that students in a given community college share some common academic and personal characteristics, this assumption is also likely to be satisfied in the current analysis. Detailed information with respect to the issue of common support between online and face-to-face takers is presented below in the "Across-School Versus Within-School Matching" section.

Results

Online Takers of Gatekeeper Courses

Among those who took English gatekeeper courses, only 8% took the courses online; for math gatekeeper courses, 7% took the courses online. For both subjects, students who took the gatekeeper courses in later semesters were more likely to take them online, which could be due to a general increase in online course offerings within VCCS across the 4-year span under study. Online sections also seemed more popular during summer terms, with 16% of summer English gatekeeper students and 14% of summer math gatekeeper students choosing online sections.

To estimate the propensity of each student to take a gatekeeper course online in each subject, we used multilevel logistic regression (see equation 2). The odds ratios of online enrollment on the basis of demographic, time, and school covariates are presented in Table 2. For both subjects, students who took the online format differed from students taking courses through the face-to-face format on multiple variables (at $p < .05$); older students, women, career-technical students, White students, English-fluent students, and students with lower credit loads in the current semester tended to have significantly higher odds of taking gatekeeper courses through the online format. Interestingly, when controlling for other baseline characteristics, students who had been dual enrolled prior to entry and students who had not previously enrolled in a remedial course in the corresponding subject area tended to have higher odds of choosing the online course format. Although these differences were not always statistically significant, the pattern may suggest that students with stronger academic preparation are more likely to choose the online format than are their academically underprepared counterparts. Thus, simple comparisons between online and face-to-face students may underestimate the negative impacts of the online format on course outcomes. In addition, the results indicate that the odds of students' attempting the online format were significantly higher during summer and in later years. For example, the odds of taking an English gatekeeper course online in the 2005–2006 academic year were more than twice as high (odds ratio = 2.15) as in 2004–2005, and the odds further increased to 2.52 in 2006–2007 and 3.60 in 2007–2008. In terms of school-level covariates (see note 8), students in colleges with lower per student instructional expenditures and larger percentages of students receiving federal financial aid had significantly higher odds of attempting a gatekeeper course through an online format.

TABLE 2
Odds Ratios of Online Enrollment in English and Math Gatekeeper Courses

	English		Math	
Pretreatment covariate	Coefficient	*SE*	Coefficient	*SE*
Demographic characteristics				
Age	1.2751***	0.2515	1.4078***	0.2838
Age^2	0.9967***	0.2051	0.9949***	0.1947
Female	1.3797***	0.3254	1.3576***	0.4466
Black (White as the base group)	0.5826***	0.1103	0.7453**	0.3618
American Indian	0.9753	16.254	0.8722	4.8457
Asian	0.6974*	0.3791	0.5377**	0.2278
Hispanic	0.5864**	0.2336	0.6762	0.4850
Academic characteristics				
Entered college at age 25 or older	1.2885	0.9912	1.0518	5.8432
Transfer track (vs. career-technical)	0.8613**	0.4328	0.6777***	0.1765
Dual enrolled prior to entry	1.2871*	0.6704	1.1055	1.6257
ESL student	0.4799***	0.1752	0.4971**	0.2511
Federal financial aid recipient	1.3433***	0.3582	1.1411	0.9202
Took remedial courses in this subject	0.9020	0.7644	0.6305***	0.1542
Full-time student in the current term	0.9829	7.0206	0.9862	12.3271
Took computer literacy course previously or concurrently	1.4047***	0.3477	1.1117	1.1461
Credits attempted in the current term	0.9609**	0.4214	0.9593*	0.5578
Timing of ENG111 enrollment				
Year 2005–2006	2.1501***	0.2995	1.6976***	0.3957
Year 2006–2007	2.5236***	0.4381	2.4653***	0.4214
Year 2007–2008	3.6020***	0.5637	2.9011***	0.5343
Summer	1.8316***	0.3685	2.1107***	0.4453
School characteristics				
Instructional expenditure per student	0.9991**	0.4803	0.9988**	0.3886
Academic expenditure per student	0.9997	5.5540	1.0014	1.3176
Student expenditure per student	0.9995	3.4467	0.9994	3.2240
Institutional expenditure per student	1.0010	0.6586	1.0010	0.7700
Percent of federal financial aid students	1.0333**	0.4806	1.0437**	0.4142
School located in rural area	0.7460	1.4920	0.7988	2.4207
School located in suburban area	0.8401	2.8000	1.1608	5.0469
Percentage of minority students**	0.9619	0.4201	0.9918	2.3065
Observations	13,951	13,951	8,328	8,328

Note. ESL = English as a second language.
*Significant at the .10 level. **Significant at the .05 level. ***Significant at the .01 level.

Across-School Versus Within-School Matching

Figure 1a presents the distributions of the probability of taking a gatekeeper course online in each subject for each group of students (those who took the course online and those who took it face-to-face). For both English and math, the distribution for the face-to-face subjects is sharply skewed to the right, with approximately 80% (n = 10,743 for English and n = 6,454 for math) having a probability of less than .10, and fewer than 1% having a probability of greater than .50 (n = 78 for English and n = 31 for math). That is, most face-to-face students had a quite low probability of taking an online course. In contrast, while the distribution for the online group is also skewed to the right, the percentage of online students having a probability of less than .10 is only about 40% (n = 396 for English and n = 234 for math), and the percentage of those having a probability of greater than .50 is about 8% for English (n = 88) and 3% for math (n =18).[9] In a well-matched sample, the two distributions should lie approximately on top of each other. In

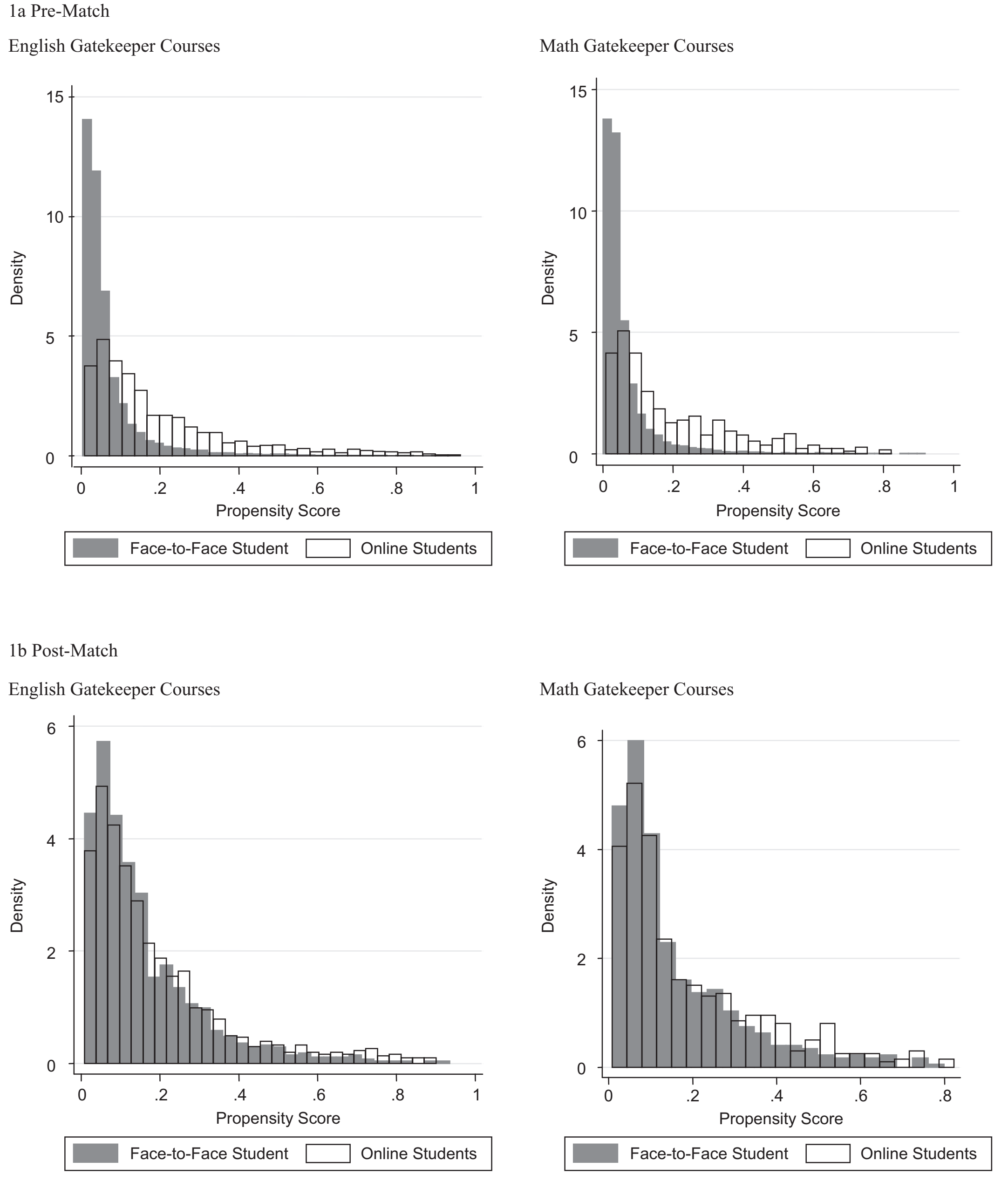

FIGURE 1. *Probability densities for the online and face-to-face course takers.*

Figure 1a, the raw samples are clearly not well matched for either English or math.

Next, we proceeded to match the samples; in the process, many face-to-face students with low probabilities of online course taking were discarded from the sample. Because the online population is much smaller than the face-to-face population in both gatekeeper course subjects, most online takers found successful face-to-face matches within 0.1 standard deviations of the propensity score. That is, very few online students were discarded, although within-school matching resulted in a slightly higher number of discarded online subjects than did across-school matching, for both English (18 vs. 1) and math (14 vs. 0). We next examined the two matching strategies in terms of covariate balance. Following Stuart (2008) and Austin (2007), we used the "standardized difference" (i.e., the absolute difference in sample means divided by an estimate of the pooled standard deviation of the variable) to check balance in group means. Some researchers (e.g. Hill, 2008) also recommend examining higher order sample balance; therefore, we also checked the ratio of standard deviations between the online group and the face-to-face group. Results for the English sample are presented in Table 1.[10] Both matching strategies resulted in satisfactory balance on all covariates in terms of both measures, although the within-school matching strategy seemed to have slightly stronger balance. For each covariate, the standardized difference in group means (a smaller number in absolute value indicates better balance) was ≤0.07 using across-school matching and ≤0.03 using within-school matching. In terms of balance on standard deviation, though the ratio of standard deviation between the online group and face-to-face group (a ratio closer to 1 indicates better balance) was ≤1.25 for both matching methods, within-school matching achieved noticeably better balance not only on all school-level variables but also on key individual covariates such as female gender (1.00 vs. 1.09), Hispanic ethnicity (1.13 vs. 1.23), and limited English proficiency (1.07 vs. 1.15).

Figure 1b shows the probability densities for the online and face-to-face students after within-school matching on the basis of the multilevel model. As is visually apparent, the matching operations achieved satisfactory overlap between the online and face-to-face students for both subjects. The sufficient overlap, together with satisfactory balance on all covariates, justifies subsequent analyses on the basis of the matching sample (n = 1,879 for English and n = 1,156 for math).

Course Attrition

On a descriptive basis, English gatekeeper courses suffered from an average attrition rate of 11% among all first-time takers, with a 9 percentage point gap between online takers (19%) and face-to-face takers (10%). The average attrition rate for math gatekeeper courses was 13%, with a 13-point gap between online takers (25%) and face-to-face takers (12%). Table 3 presents estimates of the relationship between online format and course dropout for each gatekeeper subject on the basis of a multilevel logistic model conducted with three different samples: (a) the full sample, (b) a postmatch sample constructed using across-school propensity matching, and (c) a postmatch sample using within-school propensity matching. Results from all three empirical strategies show that students who took gatekeeper courses via the online format had a significantly greater likelihood of course withdrawal for both math and English. When controlling for observed individual, time, and school variables, the odds of course dropout for English online students were 2.37 times the odds for face-to-face students (p < .001) using the full-sample specification; the coefficient falls to 2.27 (p < .001) when using across-school propensity matching. However, the within-school matching strategy further attenuated this negative impact to 1.93 (p < .001). In a similar vein, the corresponding estimates for math courses were 2.93, 2.92, and 2.70. To aid in the interpretation of each odds ratio, we also present the marginal effect, calculated by subtracting the model-predicted probability of dropout for the face-to-face course from the corresponding probability for the online course. For example, Table 3 shows that for the full sample, the percentage of students who drop out is 13.31 points higher in an online course than in a face-to-face course. This estimate is substantially higher than the raw estimate (9 points) discussed at the beginning of this section, because of the model's control for the superior background characteristics of online course takers. When we narrow the sample to compare online course takers to similar students in their own schools, the estimated gap of 10.76 points still remains slightly higher than the raw difference. The same pattern is apparent for math; the within-school matching model's estimated gap of 14.88 points is slightly higher than the raw difference of 13 points.

We also examined the changes of attrition rate in the two different course formats over time. Descriptive data imply a trend of increasing attrition rates over the 4 years, with a consistent advantage of the face-to-face format over the online format for both subjects, except in

TABLE 3
Relative Effects of Online Format on Course Outcomes

Variable	Multilevel logistic estimates based on full sample	Multilevel logistic estimates, across-school propensity score matching	Multilevel logistic estimates, within-school propensity score matching
English gatekeeper courses			
Outcome: course dropout			
Online format odds ratio (*SE*)	2.3665*** (0.2520)	2.2689*** (0.3775)	1.9311*** (0.3893)
Marginal effect	0.1331	0.1236	0.1076
Observations	13,951	1,952	1,879
Outcome: successful course performance			
Online format odds ratio (*SE*)	0.6445*** (0.1323)	0.6395*** (0.1790)	0.6700*** (0.2042)
Marginal effect	–0.0832	–0.0817	–0.0746
Observations	12,417	1,577	1,513
Math gatekeeper courses			
Outcome: course dropout			
Online format odds ratio (*SE*)	2.9359*** (0.3117)	2.9241*** (0.4059)	2.6982*** (0.4875)
Marginal effect	0.1702	0.1667	0.1488
Observations	8,328	1,156	1,128
Outcome: successful course performance			
Online format odds ratio (*SE*)	0.5898*** (0.1476)	0.5960*** (0.1523)	0.6202*** (0.1948)
Marginal effect	–0.1325	–0.1195	–0.1025
Observations	7,243	872	844

***Significant at the .01 level.

academic year 2007–2008, when online and face-to-face English gatekeeper participants had fairly equal attrition rates. To take these trends into account, we created interaction terms between year dummies and the course format to explore whether the impact of online course format on course attrition differed over the years for each subject, controlling for individual, academic, time, and school covariates. We used an F test to examine the joint statistical significance of these interaction terms using the within-school propensity score matching specification. The null hypothesis, that they were jointly insignificant, failed to be rejected ($F = 0.50$, $p = .48$ for English; $F = 0.03$, $p = .87$ for math). This result implies that the observed drop in course attrition for English online takers in 2007–2008 was driven by a shift in the composition of the characteristics of students attending online English gatekeeper courses that year. That is, the adjusted association between course format and student attrition did not change significantly over the 4-year span of the study.

Successful Course Performance

Among students who persisted to the end of the course, 77% earned a grade of C or above for English and 72% did so for math, denoting “successful” course performance. On a descriptive basis, students who took the English course online had a lower average rate of successful performance (74%) compared with those who took it in the face-to-face context (77%). The raw difference was larger for math courses; online students had a success rate of 67%, whereas face-to-face students had a success rate of 73%. We used the same three empirical strategies to explore the impact of online format on successful course performance. As Table 3 shows, all three strategies point to the same conclusion: Students who completed gatekeeper courses online had a significantly lower likelihood of earning a C or above than did those who completed face-to-face sections of the courses. When controlling for observed individual, time, and school variables, the odds of successful course performance for online students were 0.64 times the odds for face-to-face students, using the full-sample specification. The estimated odds are approximately the same using across-school propensity matching, but fall[11] to 0.67 when using multilevel within-school matching to control for unobserved school-level characteristics. The corresponding estimates for math courses were 0.59, 0.60, and 0.62. Marginal effects using the full student sample revealed a substantially larger gap (8.32 percentage points for English and 13.25 for math) compared with the raw difference (3 percentage points for English and 6 for math). Even when comparing online students with similar students in their schools, the estimated gaps remain much larger than the raw differences. This pattern of results implies that online students who persisted to the end of the course were positively selected compared with their counterparts in the face-to-face sections, and indeed the sizes of the coefficients for key academic characteristics in the propensity model for this reduced sample (n = 12,417) are slightly larger than the coefficients presented in Table 3.[12]

We also explored whether the association between online enrollment and successful course performance changed over time. Descriptive results suggest that online students consistently lagged behind their face-to-face peers in successful performance rates and, moreover, that this gap seemed to increase over time for both subjects. Although changes in the nature of online courses (e.g., increasing difficulty or deteriorating quality) could be a potential explanation, an equally compelling explanation may be that individual characteristics jointly influenced both online enrollment in later years and course outcomes. Accordingly, we further included into the post-match analytical model interaction terms between year dummies and course format to explore potential changes in the estimated treatment effect over the years controlling for individual, academic, time, and school covariates for each subject. Again, the F test failed to reject the null hypothesis that these interaction terms were jointly insignificant in predicting course performance ($F = 1.78$, $p = .18$ for English; $F = 1.78$, $p = .18$ for math). Thus, the apparent increasing gap between online and face-to-face students in course performance is a spurious pattern that can be explained by individual characteristics included in the model and is not due to changes in the nature of the online gatekeeper courses over time.

Robustness Check and Sensitivity Analysis

Because potential analytic problems may derive from propensity score model specifications, we used a range of other propensity score specifications that all yielded fairly adequate balance. Despite minor variations in the coefficients, the results on the basis of each propensity score specification were qualitatively similar to those presented in Table 1.

One potential important indicator of both online enrollment and course outcome is employment status, which is not available in the VCCS data set. However, students who applied for federal financial aid (about one third of the sample) indicated whether they were dependent on a parent, which may serve as a good proxy for the student's level of employment. Among the subset of students who provided this information (n = 5,899 for English, n = 3,702 for math), dependency status was a significant indicator of online enrollment ($p < .001$ for both subjects); however, including dependency status as an additional matching and control variable did not substantively alter the coefficient of online learning for either outcome variable.

Given the wide school-level variation in the number of gatekeeper course enrollments and in the proportion of courses taken online, we conducted a series of robustness checks to ensure that our results did not reflect the effectiveness of online gatekeeper courses in particular schools. To be specific, we reran the analyses on the basis of a sample excluding the three colleges with the largest gatekeeper enrollments in each subject, as well as on a sample excluding the three colleges with the largest online enrollments in each subject. Despite small variations, results were similar to those presented in Table 3.

Another possible problem is that unmeasured pretreatment characteristics might jointly influence both treatment status and course outcomes. Yet the omission of such variables from the current analysis would constitute the violation of the "strong ignorability" assumption only if their influences were independent of the estimated propensity scores. A sensitivity analysis following Rosenbaum's (2002) method showed that the results for course attrition would be sensitive to an upward hidden bias of $\Gamma = 2.75$ for English courses and $\Gamma = 3.50$ for math courses; that is, to question our conclusion regarding the positive association between course format and course attrition, an unobserved covariate would have to significantly increase the probability of course dropout while tripling the odds of online enrollment. In contrast, the results for the association between course format and successful performance are more sensitive to hidden bias, with a downward hidden bias of $\Gamma = 1.5$ for English and $\Gamma = 2.25$ for math. These results imply that the estimated impact of online treatment on course performance is less robust against potential unobserved covariates than is the estimated impact on attrition. The most obvious candidates for such unobserved covariates are employment and child care responsibilities. However, we controlled for these variables' effects at least in part through the inclusion of part-time enrollment status, age, and, in a further robustness check, student dependency status. Finally, results presented in both Table 1 and Table 3 indicate that online takers tend to be positively selected. If other unobserved covariates are in line with this trend, then observing and controlling for those variables would result in further magnification of the estimated negative impact of online course delivery.

To further examine this possibility, we conducted additional analyses using students' prior GPAs. Although our primary analysis included several variables to measure students' academic preparation, the fact that they were all dummy variables limited their precision. Adding GPA provides us with a good opportunity to explore the potential shift in the estimates when academic preparation is more precisely controlled. A drawback to the inclusion of GPA is that 61% of English gatekeeper students and 41% of math gatekeeper students took the given course in their first semester and were therefore subject to missing GPA values. Thus, we reconducted the propensity matching analyses[13] by further controlling for students' prior GPAs in both the propensity score estimation model and the post-match analytical model and then compared the results with estimates that used the same student sample but did not control for students' prior GPAs. Further control of prior GPA notably increased the estimated treatment effects for this subsample; the odds ratio of course dropout increased from 1.68 to 3.41 for English and from 2.58 to 3.52 for math when controlling for prior GPA, and the odds ratio of successful course performance decreased from 0.71 to 0.61 and from 0.61 to 0.57 for math. These results provide additional support for the notion that online takers tend to be positively selected and that the negative impacts of online course enrollment tend to be underestimated in the absence of key individual variables.

Conclusions

Researchers, practitioners, and policymakers are engaged in vigorous debate about the effectiveness and future promise of online learning in higher educational institutions. In an attempt to contribute reliable data on community college students' online course performance, in the current study we used a multilevel logistic regression strategy as well as two propensity matching strategies to explore the effects of taking gatekeeper courses online in the subjects of English and math. Comparisons of the across-school and within-school matching techniques demonstrated that within-school matching can achieve better balance in baseline covariates between treated and control groups. However, despite small variations in the magnitude of the treatment effects, all three empirical strategies suggest that students pay a price for taking these key introductory courses online, in terms of both course persistence and performance. Sensitivity and robustness checks suggested that the estimated negative effects of online learning are unlikely to be due to omitted variable bias and may even be underestimated, given that the inclusion of previous GPA magnified the negative effect.

Accordingly, our results strongly suggest that online instruction in key introductory college-level courses, at least as currently practiced, may not be as effective as face-to-face instruction at 2-year community colleges. Proponents of online learning may be tempted to dismiss data from 2004 as irrelevant, because online learning is purportedly evolving at a fast pace. However, our analyses showed that the estimated negative effect of online learning did not significantly change between 2004 and 2008, suggesting that evolving technologies were either not adopted or did not have a strong impact on online success rates.

Community college students face a variety of challenges as they attempt to complete courses and progress toward degrees. In addition to work and family responsibilities, these students are often academically underprepared for college-level work, and many are first-generation college students. All of these circumstances suggest that many community college students require additional instructional and institutional supports to succeed academically. Our findings suggest that online gatekeeper courses, as typically designed and implemented, do not provide these supports to community college students as effectively as do on-campus, face-to-face versions of these courses.

Proponents of online learning have consistently noted that it is not course modality but course quality that influences student learning. In principle, we agree with this assertion, but our results suggest that designing and teaching high-quality online courses with sufficient student supports may be more difficult than doing so with face-to-face courses. That is, for increased online course offerings to translate to improved academic success and postsecondary progression, institutions may need to devote substantially more resources to developing and evaluating programs and practices explicitly designed to improve such students' retention and learning in online courses. Without a more critical examination of the pedagogical factors, student supports, and institutional structures that reinforce online students' academic commitment and motivation, it is unlikely that an increase in online offerings will result in a substantial increase in educational attainment among community college students.

Notes

1. Online courses are defined here as having at least 80% of their course content delivered online and typically without face-to-face meetings (Allen & Seaman, 2005).

2. Many community colleges define successful performance in gatekeeper courses as earning a C or better, on the basis of institutional wisdom that students who do not earn at least a C are unlikely to succeed in subsequent college-level courses.

3. In the 2004 VCCS data set, distance education refers to courses with 95% or more of the content offered asynchronously. Although a few distance courses in the Virginia system are offered through television, correspondence, or other methods, almost all are offered entirely online; we will refer to these courses as "online courses" throughout the article.

4. This description is based on statistics reported to the Integrated Postsecondary Education Data System database. However, when comparing the characteristics of Virginia's community colleges with those of U.S. community colleges as a whole, none of these institutional differences reaches statistical significance at the .05 level.

5. Students were placed into remedial English and math courses on the basis of VCCS placement exams. However, placement exam scores in the current data set suffer from missing data issues for one third of the student body as well as inconsistencies arising from the use of multiple exams and are therefore not feasible to include in the inferential analyses. Student enrollment in a math or English remedial course is used in the current study to control for students' initial academic preparation in the corresponding subject area.

6. Refer to Gelman and Hill (2007) for a detailed discussion of the motivation and application of multilevel modeling.

7. The squared term for age was also added into the propensity score matching model to achieve better balance on the standard deviation for age.

8. We have four types of school-level variables: per student expenditures (instructional, academic, student, institutional), percentage of minority students, percentage of federal financial aid students, and the location of the college (rural, suburban, or urban). All are continuous variables except the location of the college.

9. Given the strong difference in the distribution of propensity scores between the treatment group and control group, "matching with placement," which allows a control group member to be used as many times as he or she is the best match (with weight adjusted later), is used by the current study to generate more precise matches.

10. We also conducted a separate check on covariate balance for math. The results were similar to those presented in Table 1.

11. In odds ratio interpretation, the impact of the treatment on the outcome variable depends on how far the odds ratio is from 1. A decrease in the odds ratio implies that the negative impact of the treatment becomes stronger after adjustment for individual covariates in the model.

12. The coefficients (in odds ratios) for the reduced sample seem slightly stronger for transfer-track (vs. career-technical) students (0.8507 for the reduced sample vs. 0.8613 for the larger sample), for federal financial aid recipients (1.3771 vs. 1.3433), and for students who took computer literacy courses previously or concurrently (1.4632 vs. 1.4047). The coefficients are quite similar (difference < 0.01) for the remaining academic variables except for dual enrollment prior to entry, for which the coefficients for the reduced sample are smaller than the coefficients for the larger sample (1.1853 vs. 1.2871).

13. Because of reduced sample size, we used only across-school matching for both models.

References

Abadie A., & Imbens G. W. (2002). *Simple and bias-corrected matching estimators for average treatment effects* (NBER Technical Working Paper No. 283). Cambridge, MA: National Bureau of Economic Research.

Adelman, C. (2005). *Moving into town—and moving on: The community college in the lives of traditional-age students*. Washington, DC: U.S. Department of Education, Office of Vocational and Adult Education.

Allen, I. E., & Seaman, J. (2005). *Growing by degrees: Online education in the United States, 2005*. Needham, MA: Sloan Consortium.

Allen, I. E., & Seaman, J. (2010). *Class differences: Online education in the United States, 2010*. Needham, MA: Sloan Consortium.

Arpino, B., & Mealli, F. (2008) *The specification of the propensity score in multilevel observational studies* (Working Paper No. 6). Milan, Italy: Carlo F. Dondena Centre for Research on Social Dynamics, Bocconi University.

Attewell, P., Lavin, D., Domina, T., & Levey, T. (2006). New evidence on college remediation. *Journal of Higher Education*, *77*, 886–924.

Austin, P. (2007). A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Statistics in Medicine*, *27*, 2037–2049.

Bailey, T., Jeong, D. W., & Cho, S.-W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review*, *29*, 255–270.

Beatty-Guenter, P. (2003). Studying distance education at community colleges. *Journal of Applied Research in the Community College*, *10*, 119–126.

Bendickson, M. M. (2004). *The impact of technology on community college students' success in remedial/developmental mathematics*. Doctoral dissertation, University of South Florida, Tampa.

Brown, W. E., Lovett, M., Bajzek, D. M., & Burnette, J. M. (2006). Improving the feedback cycle to improve learning in introductory biology using the digital dashboard. In T. Reeves & S. Yamashita (Eds.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2006* (pp. 1030–1035). Chesapeake, VA: AACE.

Calcagno, J. C., Crosta, P., Bailey, T., & Jenkins, D. (2007). Stepping stones to a degree: The impact of enrollment pathways and milestones on community college student outcomes. *Research in Higher Education*, *48*, 755–801.

Caldwell, E. R. (2006). *A comparative study of three instructional modalities in a computer programming course: Traditional instruction, web-based instruction, and online instruction*. Doctoral dissertation. Available from ProQuest Dissertations and Theses (UMI No. AAT 3227694).

Carpenter, T. G., Brown, W. L., & Hickman, R. C. (2004). Influences of online delivery on developmental writing outcomes. *Journal of Developmental Education*, *28*, 14–16, 18, 35.

Carr, S. (2000, February 11). As distance education comes of age, the challenge is keeping the students. *Chronicle of Higher Education*. Retrieved from http://www.chronicle.com

Cavus, N., & Ibrahim, D. (2007). Assessing the success rate of students using a learning management system together with a collaborative tool in Web-based teaching of programming languages. *Journal of Educational Computing Research*, *36*, 301–321.

Chambers, T. E. (2002). *Internet course student achievement: In Ohio's two-year community and technical colleges, are online courses less effective than traditional courses?* Doctoral dissertation, Bowling Green State University, Bowling Green, OH.

Choy, S. (2002). *Findings from the condition of education 2002: Nontraditional undergraduates* (Report No. NCES 2002-012). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Cohen, A. M., & Brawer, F. B. (2003). *The American community college* (4th ed.). San Francisco, CA: Jossey-Bass.

Cox, R. D. (2006). Virtual access. In T. Bailey & V. S. Morest (Eds.), *Defending the community college equity agenda* (pp. 110–131). Baltimore, MD: Johns Hopkins University Press.

Dehejia, R. & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, *84*, 151–161.

Figlio, D. N., Rush, M., & Yin, L. (2010). *Is it live or is it Internet? Experimental estimates of the effects of online instruction on student learning* (NBER

Working Paper No. 16089). Cambridge, MA: National Bureau of Economic Research.

Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Harasim, L. (1987). Teaching and learning on-line: Issues in computer-mediated graduate courses. *Canadian Journal of Educational Communications*, *16*, 117–135.

Hill, J. (2008). Discussion of research using propensity-score matching: Comments on "A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003" by Peter Austin. *Statistics in Medicine*, *27*, 2055–2061.

Howell, S. L., Laws, R. D., & Lindsay, N. K. (2004). Reevaluating course completion in distance education. *Quarterly Review of Distance Education*, *5*, 243–252.

Hyllegard, D., Heping, D., & Hunter, C. (2008). Why do students leave online courses? Attrition in community college distance learning courses. *International Journal of Instructional Media*, *35*, 429–434.

Imel, S. (1998). *Myths and realities: Distance learning*. Columbus, OH: ERIC Clearinghouse on Adult, Career, and Vocational Education.

Jaggars, S. S., & Xu, D. (2010). *Online learning in the Virginia Community College System.* New York: Columbia University, Teachers College, Community College Research Center.

Kim, J. (2006). *Causal inference in multilevel settings: Estimating and using propensity scores when treatment is implemented in nested settings*. Doctoral dissertation, University of California, Los Angeles.

Kleinman, J., & Entin, E. B. (2002). Comparison of in-class and distance-learning: Students' performance and attitudes in an introductory computer science course. *Journal of Computing Sciences in Colleges*, *17*, 206–219.

Moore, K., Bartkovich, J., Fetzner, M., & Ison, S. (2003). Success in cyberspace: Student retention in online courses. *Journal of Applied Research in the Community College*, *10*, 107–118.

Parsad, B., & Lewis, L. (2008). *Distance education at degree-granting postsecondary institutions: 2006–07* (Report No. NCES 2009–044). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Perez, S., & Foshay, R. (2002). Adding up the distance: Can developmental studies work in a distance learning environment? *THE Journal*, *29*, 16, 20–22, 24.

Peterson, C. L., & Bond, N. (2004). Online compared to face-to-face teacher preparation for learning standards-based planning skills. *Journal of Research on Technology in Education*, *36*, 345–361.

Rosenbaum, P. R. (2002) Covariance adjustment in randomized experiments and observational studies. *Statistical Science*, *17*, 286–327.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician*, *39*, 33–38.

Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association*, *95*, 573–585.

Russell, Thomas L. (2001). *The no significant difference phenomenon: A comparative research annotated bibliography on technology for distance education*. Montgomery, AL: IDECC.

Stuart, E. (2008). Developing practical recommendations for the use of propensity scores: Discussion of "A critical appraisal of propensity score matching in the medical literature between 1996 and 2003" by Peter Austin. *Statistics in Medicine*, *27*, 2062–2065.

Summerlin, J. A. (2003). *A comparison of the effectiveness of off-line Internet and traditional classroom remediation of mathematical skills*. Doctoral dissertation, Baylor University, Waco, TX.

Vargo, D. (2002). *A comparison of student success and retention rates from an intermediate algebra class presented via the internet and via a conventional classroom*. Doctoral dissertation, University of New Mexico, Albuquerque.

Authors

DI XU is a Senior Research Assistant at the Community College Research Center and a doctoral student in Economics and Education at Teachers College, Columbia University. Her current research focus includes online learning performance, remedial education, college ranking, economic returns to higher education and the use of quantitative research methods in education policy and program evaluation.

SHANNA SMITH JAGGARS is a Senior Research Associate at the Community College Research Center, Teachers College, Columbia University. Currently, Dr. Jaggars is involved in studies analyzing online course enrollment and performance; developmental education pedagogy, programming, and policy; and institutional alignment and improvement.

Manuscript received August 30, 2010
Revision received January 21, 2011
Accepted May 22, 2011