American Journal of Distance Education

ISSN: 0892-3647 (Print) 1538-9286 (Online) Journal homepage: https://www.tandfonline.com/loi/hajd20

Getting Academically Underprepared Students Ready through College Developmental Education: Does the Course Delivery Format Matter?

Bianca Cung, Di Xu, Sarah Eichhorn & Mark Warschauer

AMERICAN JOURNAL OF DISTANCE EDUCATION
2019, VOL. 33, NO. 3, 178–194
https://doi.org/10.1080/08923647.2019.1582404

Getting Academically Underprepared Students Ready through College Developmental Education: Does the Course Delivery Format Matter?

Bianca Cung[a], Di Xu[a], Sarah Eichhorn[b], and Mark Warschauer[a]

[a]University of California, Irvine; [b]University of Texas, Austin

ABSTRACT
Addressing high demand for developmental math instruction and low rates of successful completing of the developmental coursework, with cost and space constraints, has been an ongoing challenge for post-secondary institutions. With advances in online instructional technology, particularly those based on artificial intelligence, web-based instruction is increasingly considered as a way to alleviate these burdens. This is among one of the first efforts that uses a quasi-experimental design to compare the academic outcomes of students who take a developmental mathematics course in a blended setting that combines face-to-face instruction with an online intelligent tutorial system, ALEKS, to the academic outcomes of students who take the same course in a fully online setting. Results suggest that students receiving online-only instruction perform worse on the final exam and receive lower course grades. However, a cost-effectiveness analysis suggests that fully online instruction has both a lower cost per student enrolled and a lower cost per student passing the course.

Introduction

Developmental education has taken on a major role in United States postsecondary education; nationally, more than one-fifth of first-year undergraduates in public four-year universities are placed into at least one developmental course (Sparks & Malkus, 2013). The purpose of developmental education is to provide underprepared students who enter college the opportunity to strengthen their skills and bring them up to an adequate level for further college-level coursework. Mathematics is one such subject that requires students to have a core understanding of foundational topics before higher-level topics can be taught. Indeed, mathematics has the highest rate of remediation among first-year undergraduates (Attewell, Lavin, Domina, & Levey, 2006; Parsad, Lewis, & Greene, 2003) but also the lowest rate of successful developmental course completion (Bahr, 2010; Bonham & Boylan, 2011).

The large cost to institutions associated with developmental education and the high volume of students enrolled in these courses (Merisotis & Phipps, 2000; Strong American Schools, 2008) have led many institutions to transfer developmental coursework to blended or even fully online delivery format. However, students enrolled in developmental courses are often academically underprepared and may struggle particularly in a fully online learning

CONTACT Bianca Cung bcung@uci.edu
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hajd.

environment due to limited self-directed learning skills (Xu & Jaggars, 2014) and low levels of prior academic achievement (Asarta & Schmidt, 2017). While a growing volume of studies has used rigorous experimental and quasi-experimental designs to compare fully online instruction to traditional face-to-face courses or courses with varying amounts of face-to-face and online elements (e.g. Bowen, Chingos, Lack, & Nygren, 2014; Figlio, Rush, & Yin, 2010; Joyce, Crockett, Jaeger, Altindag, & O'Connell, 2015), no study has focused on the specific setting of developmental courses. Yet, understanding the relative effectiveness of different delivery formats in the specific context of developmental education is especially important given that successful completion of developmental courses plays a critical role in a student's college career, and that students who do not pass developmental courses encounter difficulty in enrolling in higher-level college courses (Calcagno, Crosta, Bailey, & Jenkins, 2007). If online courses achieve the same result as blended courses, then switching to an online instructional format is a major cost-saving strategy for universities. However, if fully online courses yield inferior results, the benefits of cost savings would be mitigated.

This study uses a quasi-experimental design to compare the academic outcomes of students who take a developmental mathematics course in a face-to-face setting with online tutorial instruction (referred to as the "blended" setting hereafter) to the academic outcomes of students who take the same developmental mathematics course in a fully online setting. This study, using administrative data from a large public four-year university, aims to uncover whether students receiving developmental math instruction through an online-only mode can perform as well as students receiving instruction of the same content through a blended mode. We take advantage of the instructional change from the blended format prior to the 2012–13 school year to the fully online format thereafter. Only one instructional mode was taught at a time, and students did not know the instructional mode of future courses, which substantially limits the likelihood of student self-selection into different delivery formats. The same set of online test banks was used to draw final exam question items across years, making a subset of questions on the final exam comparable before and after course delivery format change. We use a propensity score matching strategy to achieve balance in student baseline characteristics between different cohorts and assess the effects of taking a developmental math course online versus blended by comparing the final exam scores and course grades of students under each form of course delivery format.

Overall, our findings indicate that students achieved significantly and noticeably better final exam scores and final course grades in the blended format than in the fully online format. The results are robust to possible omitted variable bias, as indicated by a Rosenbaum sensitivity check (Rosenbaum, 2005). These results suggest that students are better supported in a developmental course with in-person lectures compared to a completely virtual learning environment. However, a cost-effectiveness analysis indicates that online-only instruction is an attractive alternative for schools with tight budgets. Administrators will need to consider both the advantages and disadvantages of online instruction when choosing between different delivery format types for developmental coursework.

Benefits and challenges of online education

In general, online instruction has been an attractive alternative to traditional instruction for numerous reasons, including its implementation as a cost-saving strategy for schools (Bartley & Golek, 2004; Clarke & Hermens, 2001) and its flexibility and convenience for students (Benbunan-Fich & Hiltz, 2003; Clinefelter & Aslanian, 2015). However, the literature on online learning suggests that online courses require students to assume greater responsibility for their learning. Success in an online learning environment is highly dependent on student initiative and self-regulation (Azevedo, Cromley, & Seibert, 2004; Corbeil, 2003; Guglielmino & Guglielmino, 2002; Hannafin & Land, 1997; Kearsley, 2002; Moore, 1987; Williams, 1996; Yen & Liu, 2009). Students who have skills in time-management and self-regulated learning are more prepared for online courses (Delen, Liew, & Willson, 2014; Pazzaglia, Clements, Lavigne, & Stafford, 2016; Wang, Shannon, & Ross, 2013; Xu & Jaggars, 2014). Students with weaker academic preparation, such as those enrolled in college developmental coursework, may also have insufficient time management and self-directed learning skills, which makes it more challenging for them to learn effectively in a fully online environment. Given the important role of developmental education at the postsecondary level and the expansion of fully online instruction in large-enrollment lower-division college courses, understanding whether delivering a developmental course completely online through well-developed ITSs, such as ALEKS, may achieve similar student learning outcomes is of great policy importance.

This study addresses this research gap by examining whether taking a developmental mathematics course in a blended setting that combines face-to-face instruction with ALEKS would lead to similar learning outcomes compared to students who take the same course in a fully online setting. Many four-year institutions have already started teaching developmental and introductory courses online; therefore, the findings of this study would have important implications for colleges and administrators who are contemplating replacing blended learning with completely online instruction in college developmental coursework.

Data and research context

Data

This study examines a natural experiment in which a developmental mathematics course "Pre-Calculus" at a major university in the western United States was taught in two different formats across three academic years from 2010 to 2013: (1) face-to-face with online tutorial homework (blended) and (2) fully online with no face-to-face instruction. In the 2010–12 academic years, Pre-Calculus was taught in a blended setting with online tutorial instruction in addition to four hours of face-to-face meetings each week. In the 2012–13 academic year, Pre-Calculus was taught fully online with no scheduled face-to-face meetings. Across all three years, students worked individually online using the same online tutorial program.

The data for this study came from two sources: the university's Office of Institutional Research (OIR) and the online adaptive tutorial program, ALEKS. The OIR dataset included information about students' academic and demographic background, such as students' grade in the course, SAT scores, major, year admitted to the university, gender, and ethnicity. The ALEKS dataset contained students' tutorial completion and exam scores. Students had a series of short tutorials to complete by a pre-determined course deadline. Responses to the short tutorials allow the ALEKS to determine whether the student mastered the content and what topics the student was prepared to learn next. Students could pick a topic, from a list suggested by ALEKS, that they would like to learn next. ALEKS determined the initial list of topics that students were ready to learn through an initial assessment test (an adaptive questioning system). The initial assessment and final exam scores are recorded. The midterm exam included 30 question items, and the final exam included 40 question items with topics pre-determined by the school's mathematics department. The students individually had different exam questions randomly drawn from ALEKS's test bank according to each topic.

A total of 2,117 students enrolled in the Pre-Calculus course during fall and winter quarters across the three academic school years. We limited the analysis sample to 1,528 non-transfer students who had never taken Pre-Calculus previously at the university. We also excluded 157 students who did not complete the ALEKS final exam. The proportion of students skipping the final exam is similar for both the blended (13%) and fully online (10%) formats. A total of 27 ALEKS accounts could not be successfully matched to students in the OIR dataset, and they were excluded from the study. We additionally excluded eight students who had previously scored passing grades (3 or higher) on the AP Calculus AB or BC exams, one student missing an SAT Math score, and three students missing gender information in the OIR dataset. The final analysis sample consisted of 1,332 students. Enrollment counts by term can be found in Table 1.

Course format description

Blended

During the 2010–11 and 2011–12 academic years, Pre-Calculus was taught over the full ten-week school term in a blended format with online intelligent tutorial instruction in addition to four hours of weekly face-to-face time. These two years were very similar in format and support. The course was taught in sections of approximately 120 to 130 students. Students were expected to meet in these large sections three times each week for an hour each time. Class meeting time was scheduled for students to meet with a teaching assistant in smaller sessions, called discussion sections. Discussion sections met in classrooms of approximately 60 students for one hour each week.

The courses in each term of those two years used the same textbook, offered support through in-person office hours, and administered quizzes during the discussion sections. However, the grading scheme and number of ALEKS benchmarks differed in the first and second years. Benchmarks were deadlines for a certain set of topics on ALEKS. Although the same topics were covered for each course on ALEKS across the terms, the 2010–11 year had three benchmarks while the 2011–12 year had four benchmarks. More lessons were due, on average, at each benchmark in the 2010–11 year since there were fewer benchmarks.

Table 1. ALEKS enrollment by term.

	Instructional Mode	Total	Analysis Sample
Fall 2010	Blended	361	269
Winter 2011	Blended	237	142
Fall 2011	Blended	329	221
Winter 2012	Blended	331	195
Fall 2012	Online Only	500	328
Winter 2013	Online Only	359	177
Total		2,117	1,332

Grades in fall 2010 and winter 2011 were determined by three ALEKS benchmarks (worth 5% each), an ALEKS final exam (worth 5%), quizzes administered during the discussion section (worth 10%), two paper midterm exams (worth 20% each), and a paper final exam (worth 30%). Grades in fall 2011 and winter 2012 were determined by four ALEKS benchmarks (worth 5% each), an ALEKS final (worth 5%), discussion quizzes (worth 10%), a single paper midterm exam (worth 25%), and a paper final exam (worth 40%). The math course examined in this study, for the most part, followed the recommendations suggested by the ALEKS creators in that a portion of students' final grades were contingent on work done on ALEKS, coursework assigned in a traditional course was reduced to account for additional work done on ALEKS, and student progress was checked throughout the term (every other week instead of weekly).

Students in both years took both a paper-based final exam, as well as a computer-based final exam on ALEKS, though final grade weights given to the paper final exams changed in the second year as the number of midterm exams changed. The first year's sections administered two paper midterm exams while the second year administered only one. Courses also had benchmarks representing deadlines when students had to complete ALEKS lessons on a list of topics. ALEKS's role in the course changed from three benchmarks and one online final exam in fall 2010 to four benchmarks and one online final exam in winter 2011. The ALEKS benchmarks and final added up to 20% of the student's grade. In the 2011–12 school year, the courses had a 20% weight on four benchmarks and an additional 5% weight on a final exam taken on ALEKS. Overall, the grading scheme was consistent within the same academic year but changed slightly from the first year of blended offerings to the second.

In a typical week, students in the blended course attended face-to-face lectures and discussion sections on campus during regularly scheduled times. Lessons were delivered at an instructor-determined pace. At home, students worked on lessons on ALEKS at their own pace. Students were expected to complete the lessons before each benchmark deadline. Lessons on ALEKS were determined by each student's initial test and their progress solving lesson-related problems. Since ALEKS was completed at the students' pace, it was possible for face-to-face lectures and online ALEKS work to be unaligned.

Online

The online-only format was offered in the 2012–13 academic year, using the same textbook as in the blended format. Students did not receive in-person lectures but instead worked individually through the ALEKS instructional system over the full ten-week school term. Instructors were available at specified hours for both in-person and online consultations. Each section had one midterm and one final exam, both of which were weighted more heavily compared to other assignments. In both the fall and winter terms, the midterm and final exams were taken on ALEKS. In fall 2012, the course was divided into four benchmarks, each worth 7% of the final grade. The remaining 72% was based on students' performance on attending one office hour during the first two weeks of class (worth 2%), one midterm exam (worth 30%) and one final exam (worth 40%). In winter 2013, the course was divided into six benchmarks, each worth 4% of the final grade.

Additionally, the students who took Pre-Calculus in winter 2013 had four quizzes, each worth 2% of their final grade. The remaining 68% was based on students' performance on an orientation quiz (worth 3%), a midterm exam (worth 25%), and a final exam (worth 40%). The orientation quiz aimed to help students better understand the format and expectations of the course. Both the midterm and final exams were proctored and taken on ALEKS. With the exception of shorter intervals between benchmarks and the incorporation of online quizzes, the winter 2013 Pre-Calculus offering was very similar to that of fall 2012.

In a typical week, students would work on lessons on ALEKS at their own pace. Like the blended group, they would complete lessons before each benchmark deadline. However, since there were no face-to-face lectures, students went over course topics at their own pace.

Although the course structure changed from term to term, one unique attribute of this data set is that only one course format was offered at a time. Student self-selection into course formats would have been difficult since students did not know what format would be offered in the future and since mathematics was a pre-requisite to many STEM courses. To address possible variations between cohorts in student characteristics, we employ a propensity score matching strategy, which we explain in detail in the following sections.

Methodology

Outcome measures

Student performance was measured by ALEKS final exam scores and course grades. We focus primarily on a modified final exam score, which we will call the final exam sub-score, as it only includes question items drawn from the same test bank across all three years. Though not presented here, the results of the raw percent final exam score yielded similar results to that of the modified final exam score.

The final exam sub-score, represented as a percent, is calculated from specific question topics and types asked across all three years. If all the six terms across the three years asked questions drawn from the same item bank (such as "Double-angle identities: Problem type 1"), then students' scores corresponding questions from that particular item bank were included in the final exam sub-score calculation. Questions on the same topic but different item banks ("problem type 1" versus "problem type 2") were not included, as one item bank may have been more difficult than the other. A total of 11 question topics, covering a range of different course content, are included in the calculation of the final exam sub-score, whereas the full final exam contained 40 question items each.

In addition to the final exam sub-score and final exam score, we also include results for course grades. We convert to a 4.0 numeric scale. The grades "B-" and "B+," for example, equate to 2.7 and 3.3 grade points respectively. The grade "F" equates to 0 grade points. We set "A+" as 4.3 grade points to distinguish the higher-valued A-grade, despite how some schools calculate both "A+" and "A" as 4.0 grade points. While student GPA calculations do not reflect "pass" grades in school records, we convert "pass" to 2 (the equivalent of a C) and "no pass" to 0. We include course grades because it is an important reflection of student success in a course. However, results need to be interpreted cautiously, as student grade points for the course may be influenced by different assignment weights across terms.

Propensity score matching

As student background and characteristics vary from term to term, it would be inappropriate to draw a direct comparison between all students in both conditions. This is especially true for the sample used in this study since the university changed some of its admission criteria, placing less emphasis on standardized exam scores, just before Pre-Calculus was offered in an online-only format. Along with the admission change, some individual departments admitted students with lower levels of preparation in later years. We use propensity score matching (PSM) to generate two comparable groups of students who would have been just as likely to have received the instruction in the counter condition if it were not for timing.

PSM is a two-step process which involves first mapping a series of covariates onto a unidimensional value (a propensity score) that represents the probability of being in the treatment group (defined as receiving the instruction through online-only delivery format). For this first step, we use logistic regression to calculate each individual's propensity score. In the second step, propensity scores for each person in the two groups are matched and re-weighted so that the treatment and control groups are similar. We use Mahalanobis distances and kernel matching with a bandwidth of 1.5 and with replacement. For reference, propensity score kernel smoothings showing common support and match distributions can be found in Figure 1. The results meet the following balance criteria: a standardized difference of at most 0.1 and a standard deviation ratio between 0.95 and 1.10 (Table 2). Balance checks for each individual term can be found in the appendix. As a sensitivity check, we use bandwidths of 1 and 2 to include different numbers of students on support and check the consistency of our results. By matching students in the blended condition to students in the online-only condition, we select a sample of students who closely matches the distribution of student characteristics in the online-only year. We also conduct a sensitivity check by matching students in the online-only condition to the blended condition; the results are virtually the same.

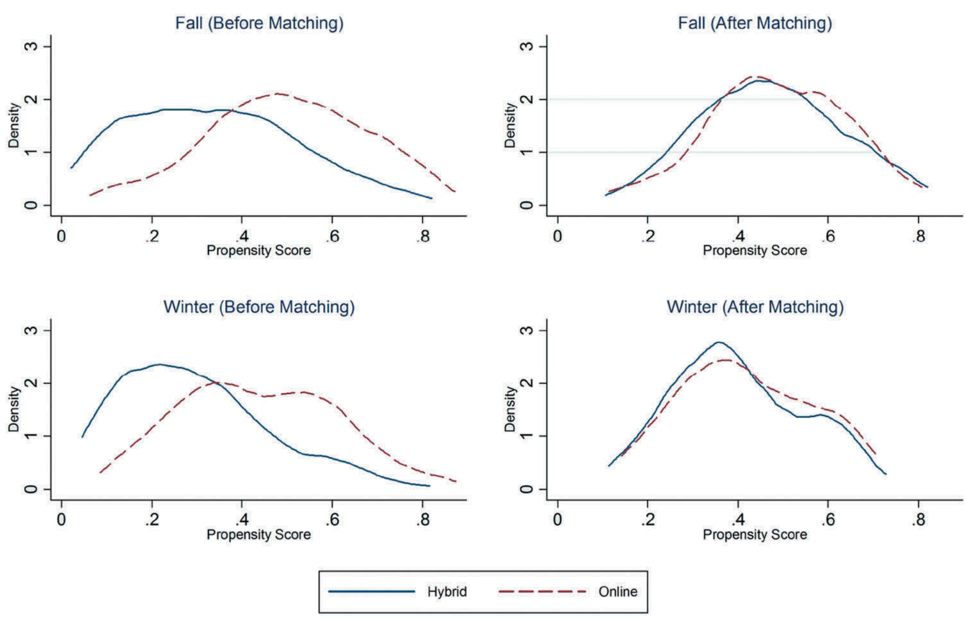

Figure 1. Propensity score distribution before and after matching with weights.

Table 2. Pooled PSM balance check for fall and winter students.

		Mean			Standard Deviation		
Variable	Match[a]	Blended	Online	Std. Diff.[b]	Blended	Online	Ratio[c]
Initial Assessment	Pre-	35.82	24.19	−0.73	22.62	15.84	0.70
	Post-	24.47	24.07	−0.03	12.82	12.67	0.99
SAT Math Score[d]	Pre-	−0.46	−0.80	−0.48	0.70	0.72	1.02
	Post-	−0.72	−0.73	−0.02	0.54	0.53	0.98
SAT Verbal Score[d]	Pre-	−0.63	−1.05	−0.50	0.88	0.85	0.97
	Post-	−0.92	−0.96	−0.04	0.70	0.72	1.02
Female	Pre-	0.62	0.61	−0.03	0.49	0.49	1.01
	Post-	0.64	0.64	0.00	0.48	0.48	1.00
Ethnicity							
Asian	Pre-	0.41	0.31	−0.22	0.49	0.46	0.94
	Post-	0.38	0.38	0.00	0.48	0.48	1.00
Black	Pre-	0.03	0.05	0.08	0.17	0.21	1.22
	Post-	0.01	0.01	0.00	0.09	0.09	1.00
Hispanic	Pre-	0.35	0.45	0.20	0.48	0.50	1.04
	Post-	0.52	0.52	0.00	0.50	0.50	1.00
White	Pre-	0.15	0.11	−0.13	0.36	0.32	0.88
	Post-	0.09	0.09	0.00	0.29	0.29	1.00
Other	Pre-	0.05	0.08	0.11	0.22	0.27	1.23
	Post-	0.01	0.01	0.00	0.08	0.08	1.00
Major Requires Math	Pre-	0.62	0.58	−0.09	0.48	0.49	1.02
	Post-	0.62	0.62	0.00	0.49	0.49	1.00
Freshman	Pre-	0.70	0.78	0.19	0.46	0.42	0.91
	Post-	0.82	0.82	0.00	0.39	0.39	1.00

Note. Students in the treatment group received instruction fully online while students in the control group received blended instruction, both face-to-face and online.

[a] "Pre-" reflects the unmatched estimates, and "post-" reflects the estimates after matching from PSM. PSM was done using Mahalanobis distances and kernel matching with a bandwidth of 1.5.

[b] Standardized mean differences, calculated from the following formula from Austin (2008):

$$SD = \frac{\left|\bar{X}_{online} - \bar{X}_{hybrid}\right|}{\sqrt{\frac{S^2_{online}+S^2_{hybrid}}{2}}}$$

[c] The ratio of the standard deviation between the online-only group and the blended group.

[d] SAT Math and Verbal scores are centered at a score of 600 and have been divided by 100.

Covariates for matching include initial assessment score, SAT math and verbal scores, gender, ethnicity, whether the student has declared a major that requires math courses, and whether the student is in his or her first year at the university. Since the purpose of propensity score matching is to achieve balance in the entire distribution of baseline covariates rather than just means, some researchers (e.g., Austin, 2011) suggest adding nonlinear terms of covariates to help achieve balance in standard deviation. Therefore, we also added the square root of the initial assessment score into the matching model in addition to other variables for individual background characteristics. As there may be differences in student characteristics based on the term that they take Pre-Calculus (e.g., weaker students may hold off on taking Pre-Calculus until winter), we match fall students to fall students and winter students to winter students.

To assess the possibility that our models still face the threat of missing variable bias, we further use Rosenbaum's sensitivity analysis (Rosenbaum, 2005) to check the magnitude of the missing variable needed to overthrow our results.

Results

Validity of the propensity score matching strategy

Self-selection and withdrawal pose a threat to the PSM strategy used in this paper. While we were unable to account for withdrawal rates prior to the two-week drop deadline, we were able to account for withdrawals after the drop deadline. Our results suggest that only five first-time non-transfer enrollees in the blended course and three first-time non-transfer enrollees in the online-only sample withdrew from the course after the drop deadline. There was no significant difference ($p = 0.64$) between the withdrawal rates in the two course formats.

Results showed that students in the online-only and blended groups had many disparities prior to matching, partly due to the admission policy change. The two groups initially differed in SAT Math scores by about 34 points ($d = 0.46$) and SAT Verbal scores by about 42 points ($d = 0.47$), with the blended group having higher scores (Table 2). This result is a not surprising due to changes to the admission criteria. The blended and online-only groups additionally differed on their initial assessment scores with an average difference of about 12 percentage points ($d = 0.53$), also with the blended group receiving higher scores. Without matching, the treatment and control groups differed on various background characteristics. The blended instructional mode had about 10% more Asian students than the online-only mode whereas the online-only instructional mode had about 10% more Hispanic students than the blended mode. A higher proportion of the students in the fully online group were freshmen. A lower proportion of the students in the fully online group were in majors that required math courses.

Students in the blended and fully online groups had a starkly different propensity score distribution for being in the treatment condition (Figure 1). Nevertheless, we were able to identify a group of students from the blended years with similar propensity scores as students in the online-only year. After applying the appropriate weights calculated from PSM, we found a comparable sub-group of students with similar baseline characteristics on average.

Estimated impacts on course outcomes

Based on the post-match sample, we examined the treatment effects after controlling for baseline characteristics. The first columns in Tables 3 and 4 present the estimated impacts of taking Pre-Calculus through the online-only format versus blended format on the final exam sub-score and course grade points, respectively, without controlling for any student-level characteristics. Overall, students who received blended instruction outperformed students who received online-only instruction across all three outcome measures: they differed by more than 10 percentage points on the final exam sub-score, and by more than half of a letter grade for the final course grade (such as from B- to B+). The results are virtually the same after controlling for students' background characteristics (column 2).

Considering that the treatment effects might differ by fall and winter quarter due to changes in student population taking Pre-Calculus, columns 3 and 4 in Tables 3 and 4 present the estimated impacts of online-only format for fall and winter quarters separately, and column 5 used the same sample and model as in column 2 – but further included an interaction term between the treatment (online-only) and whether the course was taken

Table 3. Treatment effect on final exam sub-score.

	Model 1	Model 2	Model 3	Model 4	Model 5
	Pooled[a]	Pooled[a]	Fall Only	Winter Only	Pooled[a]
Online	−11.19***	−10.88***	−11.32***	−9.83***	−11.34***
	(1.71)	(1.60)	(2.00)	(2.64)	(1.99)
Winter		2.29			1.57
		(1.73)			(2.59)
Online × Winter					1.42
					(3.27)
Initial Assessment		−1.88***	−1.94***	−1.82**	−1.88***
		(0.36)	(0.48)	(0.58)	(0.36)
Sqrt(Initial Assessment)		21.60***	21.68***	21.91***	21.61***
		(4.08)	(5.41)	(6.26)	(4.09)
SAT Math Score[b]		4.87*	5.06+	5.04	4.88*
		(1.96)	(2.61)	(3.19)	(1.96)
SAT Verbal Score[b]		1.45	1.82	0.49	1.43
		(1.25)	(1.51)	(2.39)	(1.25)
Female		0.80	−0.29	3.75	0.80
		(1.80)	(2.18)	(3.55)	(1.80)
Ethnicity[c]					
Asian		1.45	0.83	3.51	1.44
		(2.72)	(3.24)	(4.02)	(2.73)
Black		−3.05	−3.09		−3.06
		(9.96)	(10.14)		(10.01)
Hispanic		−1.89	−1.85	−1.34	−1.89
		(2.67)	(3.17)	(3.88)	(2.67)
Other		16.95***	15.47***	14.44	16.92***
		(3.87)	(3.70)	(9.65)	(3.84)
Major Requires Math		−0.70	−0.33	−2.51	−0.70
		(1.68)	(2.13)	(3.24)	(1.67)
Freshman		−3.08	−2.62	−5.76	−3.08
		(2.09)	(2.45)	(4.18)	(2.09)
Constant	73.44***	22.59*	24.56	20.72	22.78*
	(1.29)	(11.49)	(15.33)	(17.18)	(11.57)
N	709	709	475	234	709

Note. Standard errors in parentheses. Models include weights from propensity score matching.
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
[a] Models 1, 2, and 5 pool both fall and winter samples.
[b] SAT Math and Verbal scores are centered at a score of 600 and have been divided by 100.
[c] "White" is the base case for ethnicity.

during the winter quarter. Results suggest that the gap between the online-only and blended groups is smaller among students enrolled in winter quarters than fall quarters across all three outcome measures, though such difference does not reach statistical significance.

Sensitivity checks

We conducted several robustness checks to assess whether the results might be sensitive to different matching strategies and omitted variable bias. First, we conducted the analysis again with matching using kernel bandwidths of 1 and 2. We also reversed the matches such that blended students were selected to match the sample of online students. Table 5 shows results on final exam sub-score with different matching strategies and the results are fairly consistent with those presented in Table 3.

Despite consistency in results after choosing different bandwidths for matching, PSM faces the threat of omitted variable bias. We address this by using Rosenbaum's sensitivity

Table 4. Treatment effect on grade points.

	Model 1	Model 2	Model 3	Model 4	Model 5
	Pooled[a]	Pooled[a]	Fall Only	Winter Only	Pooled[a]
Online	−0.65***	−0.63***	−0.71***	−0.45**	−0.71***
	(0.11)	(0.10)	(0.13)	(0.15)	(0.13)
Winter		0.20+			0.07
		(0.11)			(0.17)
Online × Winter					0.25
					(0.20)
Initial Assessment		−0.11***	−0.11***	−0.13***	−0.11***
		(0.02)	(0.03)	(0.03)	(0.02)
Sqrt(Initial Assessment)		1.38***	1.33***	1.58***	1.38***
		(0.23)	(0.33)	(0.32)	(0.23)
SAT Math Score[b]		0.51***	0.48**	0.55**	0.51***
		(0.12)	(0.16)	(0.19)	(0.12)
SAT Verbal Score[b]		0.13	0.18+	−0.02	0.13
		(0.09)	(0.11)	(0.15)	(0.09)
Female		0.10	0.06	0.29	0.10
		(0.11)	(0.14)	(0.21)	(0.11)
Ethnicity[c]					
Asian		0.16	0.16	0.41	0.16
		(0.15)	(0.18)	(0.34)	(0.15)
Black		0.64	0.64		0.64
		(0.46)	(0.46)		(0.47)
Hispanic		0.05	0.03	0.21	0.05
		(0.16)	(0.18)	(0.35)	(0.16)
Other		0.53	−0.51	1.82**	0.52
		(0.60)	(0.45)	(0.63)	(0.59)
Major Requires Math		−0.03	0.13	−0.40*	−0.03
		(0.11)	(0.14)	(0.17)	(0.11)
Freshman		−0.01	0.03	−0.13	−0.01
		(0.14)	(0.16)	(0.30)	(0.14)
Constant	2.33***	−1.34*	−1.22	−1.83+	−1.30*
	(0.09)	(0.61)	(0.85)	(0.94)	(0.62)
N	709	709	475	234	709

Note. Standard errors in parentheses. Grade points are calculated on a 4.0 scale. Models include weights from propensity score matching. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[a] Models 1, 2, and 5 pool both fall and winter samples.

[b] SAT Math and Verbal scores are centered at a score of 600 and have been divided by 100.

[c] "White" is the base case for ethnicity.

analysis (Rosenbaum, 2005) to estimate how large the omitted variable bias needs to be to overthrow our results. For the fall matches, the confidence interval includes 0 when gamma, a measure of bias needed to overthrow results, reaches a value of 4.5 for final exam sub-score. The gamma values are even higher for final exam score and grade points. It means that an unobserved factor would have to increase the odds of receiving online-only instruction by an omitted factor of 4.5, after controlling for all observable student individual characteristics, in order to overthrow results for final exam sub-score in the fall. For the winter matches, the confidence interval includes 0 when a gamma reaches a value of 5.5 for final exam sub-score. That is, an unobserved factor would have to increase the odds of receiving online-only instruction by an even greater factor in order to overthrow results for the other outcome variables. Both of the gamma values are extremely high, indicating that the estimated results are unlikely to be overthrown due to an omitted variable (see Rosenbaum, 2005 for a more detailed explanation of this sensitivity check).

Table 5. Sensitivity check: Treatment effect on final exam sub-score using different matches.

	Unmatched[a]	BW(1)[b]	BW(2)[b]	Reversed[c]
Online	−10.57***	−10.16***	−10.05***	−8.25***
	(1.12)	(1.77)	(1.47)	(1.43)
Winter	2.86**	2.24	2.61	1.70
	(1.05)	(1.92)	(1.59)	(1.57)
Initial Assessment	−0.65***	−2.00***	−1.61***	−1.50***
	(0.14)	(0.39)	(0.32)	(0.39)
Sqrt(Initial Assessment)	9.66***	22.87***	18.95***	17.95***
	(1.63)	(4.44)	(3.65)	(4.46)
SAT Math Score[d]	5.82***	4.56*	5.04**	2.94+
	(0.87)	(2.25)	(1.66)	(1.78)
SAT Verbal Score[d]	−0.26	1.49	1.09	2.06+
	(0.67)	(1.44)	(1.06)	(1.17)
Female	−0.10	0.59	0.35	0.68
	(1.05)	(2.06)	(1.59)	(1.66)
Ethnicity[e]				
Asian	−1.26	0.27	0.73	−1.10
	(1.59)	(2.84)	(2.58)	(2.25)
Black	−1.90	4.38	−6.25	−6.39
	(2.93)	(3.83)	(6.53)	(8.48)
Hispanic	−2.50	−3.94	−1.35	−3.73+
	(1.63)	(2.72)	(2.59)	(2.21)
Other	−2.79	16.01***	12.32***	13.46***
	(2.53)	(4.09)	(3.33)	(3.44)
Major Requires Math	−0.39	−1.15	0.05	−2.28
	(1.04)	(1.90)	(1.54)	(1.50)
Freshman	−0.45	−1.50	−2.56	0.31
	(1.17)	(2.46)	(1.99)	(1.95)
Constant	48.13***	19.56	27.95**	29.55*
	(4.88)	(12.58)	(10.27)	(12.44)
N	1,332	584	820	709

Note. Standard errors in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
[a] The "unmatched" model includes all students in the analysis sample and does not use PSM.
[b] "BW(1)" and "BW(2)" models use Mahalanobis and kernel PSM with bandwidths of 1 and 2, respectively.
[c] The "reversed" model matches the blended years to the online year using Mahalanobis and kernel PSM with bandwidth 1.5.
[d] SAT Math and Verbal scores are centered at a score of 600 and have been divided by 100.
[e] "White" is the base case for ethnicity.

Alternative explanations for the gaps

One potential reason why a performance gap exists between the online-only course and the blended course could be the fact that students in the online-only course spent less time studying the content matter. After re-weighting from the results of the PSM step, the average time that blended students spent on ALEKS was approximately 66 hours. In contrast, the average time that online-only students spent on ALEKS was approximately 89 hours. Online-only students spent 23 more hours studying with ALEKS than blended students. However, blended students met face-to-face for four hours each week for ten weeks, totaling to 40 hours in-person. While we were unable to account for whether blended students actually attended in-person lectures and discussions, nor were we able to account for additional time spent studying the subject matter outside of the online environment, it is possible that advantages to blended instruction were due in part to more total time on task.

A second possible factor for why a performance gap exists between the two instructional modes could be the fact that the 2012–13 year was the only year that Pre-Calculus was taught in the online-only mode. As a result, the online-only scores may be an underestimate due to an implementation dip (Fullan, 2014). Unfortunately, further analysis of an implementation dip was not possible as the subsequent year did not offer the course in an online-only format in view of the performance dip during the 2012–2013 academic year. The possibility of an implementation dip is thus a limitation of this study. However, it is worth noting that the university had been offering ALEKS online for two years as part of the previous blended course. Nothing new was added to the course; the face-to-face element was simply eliminated.

Cost-effectiveness of blended versus online instruction

Despite overall lower outcomes, online instruction is attractive because it serves as a cost-saving strategy for postsecondary institutions. Therefore, we provide a simple cost-effectiveness analysis to compare the costs of different course delivery methods. In our cost estimates, we used an itemized costs analysis that takes into account costs related to instructors, teaching assistants, classroom space, and course development cost. The university had a coordinator for the online courses foreseeing that administration as management would be more demanding in an online course without teaching assistants. In our calculations, we included the cost of a coordinator whose role was to manage the logistics and administrative details of online courses. For purposes of calculation, we estimated the total cost for a cohort of 480 students, which is the approximate fall quarter enrollment count in the course in 2012 and the years after. A cohort of 480 students equates to four blended courses or two fully online courses.

In Table 6, we have outlined and totaled the projected cost of instructing 480 students through the blended and fully online course delivery format. Blended instruction would cost the school a total of $59,600, while fully online instruction would cost the school a total of $25,485 (Table 6). The cost per student in a blended course is approximately $124. The cost per student in an online-only course is approximately $53.

To better compare blended and online courses, we take both costs and learning outcomes into account and examine the cost-effectiveness of each instructional format. Specifically, we take into consideration the overall cost and the pass rate to calculate the cost per passing student in both mediums. We adopt a similar cost-effectiveness calculation method as McEwan (2012) and Proffitt (2014). Since we are using projected estimates for 480 students, we divide the total cost by the expected number of students earning at

Table 6. Total cost of blended and online instruction.

	Blended		Online-Only	
Instructor	4 at $8,861/instructor	$35,444	2 at $8,861/instructor	$17,722
Teaching Assistant	4 at $5,900/TA	$23,600	None	$0
Classroom Space[a]	4 at 4 hours/week	$36,000	None	$0
Development Cost	None	$0	$30,000 amortized over 9 quarter offerings	$3,333
Course Coordinator[b]	None	$0	1 at $4,430/coordinator	$4,430
Total		$59,600		$25,485

Note. Cost estimates assume cohort sizes of 480, with four sections of 120 students in the blended instructional mode and two sections of 240 in the online-only instructional mode.

[a] Classroom space is calculated for four course sections using classroom space for three one-hour lectures (at $250 per hour) and one one-hour discussion (at $150 per hour) per week over ten weeks.

[b] Course coordinators are instructors who get a one course release per year (assuming 9 courses per year) to coordinate three quarters of Pre-Calculus

Table 7. Cost-effect ratio of blended and online instruction.

	Total Cost	Course Mastery	Percent Passing[a]	Cost-Effect Ratio[b]
Blended	$59,600	Final Exam Sub-Score	67.12	$185.99
		Course Grade	71.82	$172.89
Online-Only	$25,485	Final Exam Sub-Score	37.46	$141.73
		Course Grade	49.24	$107.83

Note. Cost estimates assume cohort sizes of 480, with four sections of 120 students in the blended instructional mode and two sections of 240 in the online-only instructional mode. We defined students passing the final sub-exam as students earning at least 70% or better on their final exam sub-score. Students with course grades of at least a C or better passed the course.

[a] Percent passing reflects the sample that passed either the final exam sub-score or earning passing grades in the course after matching and re-weighting through PSM.

[b] The cost-effect ratio is calculated by dividing total cost by expected number of students passing out of a cohort of 480 students. To get the expect number of students passing, we multiplied the percent passing by 480.

least 70% or higher on the final exam sub-score. Since final course grades were awarded based on different weighting scales (and therefore not fully comparable), we re-calculate the cost-effectiveness ratio using pass rate to check for consistency (Table 7).

The cost-effectiveness ratio of the blended course was approximately $186 per passing student (on the final exam). For the online course, it was approximately $142. Although more students in the blended course earned final exam sub-scores of 70% or higher, the cost per student passing the final sub-exam ended up being higher than the online-only case. If a school can successfully identify students who will not do well in the online-only course, the extra savings from the online course could go into support structures (e.g., face-to-face discussion sections, computer lab requirement) targeting underperforming students. Alternatively, offering the online course only to students who are prepared to succeed in an online-only environment can, in and of itself, serve as a cost-saving measure.

Discussion

A considerable body of prior research suggests that blended instruction has advantages compared to online-only instruction (Means, Toyama, Murphy, Bakia, & Jones, 2009). The findings of this study are consistent with existing literature. However, comparative research has not been conducted on blended versus fully online use of an ITS for teaching a full-length college course. This is an important field to study since many believe that ITSs are a highly effective form of instruction. Our study found that even when using a well-developed ITS, learning gains are still significantly greater when combined with in-person lectures. Postsecondary institutions using ITSs for developmental mathematics instruction will therefore want to weigh carefully the added educational value of face-to-face lectures. Schools that are unable to offer regular face-to-face lectures alongside web-based tutorial instruction, perhaps due to limited resources or other constraints, may want to consider offering alternative forms of support that could mimic face-to-face lectures (such as targeted small group discussions or live Web lectures). Specifically, such alternative forms of support should target students' self-regulated learning and time management as other studies of have these skills to be markers of online course readiness (Delen et al., 2014; Pazzaglia et al., 2016; Wang et al., 2013; Xu & Jaggars, 2014).

Despite the less desirable student outcomes found in this study, online-only instruction is still an attractive option due to the lower costs to the institution, as found in our cost-effectiveness analysis. Lower rates of success in online courses could instead be the result of students spending less time on the subject matter. If it were the case that time-on-task was the contributing factor to the performance gap, then requiring a certain number of hours of work weekly, either at home or in a computer lab, could help students do better in a virtual course. Additional studies should examine time-on-task in an online course. Investigations should look into whether a weekly time requirement can achieve the same results as students attending lectures for an equivalent amount of time.

Our results need to be interpreted with caution due to several limitations. First, our study only draws from a sample covering two years of blended instruction and one year of fully online instruction. Interpretation needs to take into consideration the changing course elements, as well as the possibility of an implementation dip (Fullan, 2014) from using only the first year of online-only instruction. The six different school terms included in this study had some changes to course structure over time with different numbers of exams, varying exam weights, and an increasing number of benchmarks. This study considered the elements (e.g., quizzes, homework, office hours) of each course offering to be part of the overall blended or online-only package. Due to serial correlation, we were unable to tease out which small elements contributed to differences in outcomes. These smaller course components need to be considered in future studies.

The blended courses in this study put less weight on the ALEKS benchmarks and the ALEKS final exam, whereas the online-only courses put a large proportion of students' grade on the ALEKS final exam. Even with less weight on the ALEKS final exam, students in the blended condition outperformed students in the online-only condition. If different weights on the ALEKS final exam did have an effect, it is likely that students in the blended condition paid less attention to the ALEKS final exam (due to its smaller impact on their course grade). The results from final exam scores in this study would instead be an underestimate of the actual gap between blended and online-only students.

One final limitation of the study is that it took place in a particular context: one university deploying one ITS. ALEKS has been well studied and generally found to be effective, but we cannot necessarily assume that the same results would be found for other ITSs. Also, while the developmental population at this university may be similar to developmental students at other four-year universities in the United States, they may be quite different from developmental students in other settings, such as community colleges.

Despite these limitations, however, this study takes an important step toward understanding the impacts of using a fully online instructional format in college developmental courses; it further provides the first cost-effectiveness analyses of online-only instruction versus blended learning in a full-term college course.

Acknowledgments

This paper is based upon work supported by the National Science Foundation under Grant Number 1535300. We would also like to thank Dr. Rachel Baker and Dr. Jade Jenkins for their valuable support and feedback in this study.

References

Asarta, C. J., & Schmidt, J. R. (2017). Comparing student performance in blended and traditional courses: Does prior academic achievement matter? *The Internet and Higher Education*, *32*, 29–38. doi:10.1016/j.iheduc.2016.08.002

Attewell, P., Lavin, D., Domina, T., & Levey, T. (2006). New evidence on college remediation. *Journal of Higher Education*, *77*, 886–924. doi:10.1353/jhe.2006.0037

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*(3), 399–424. doi:10.1080/00273171.2011.568786

Azevedo, R., Cromley, J. G., & Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology*, *29*(3), 344–370. doi:10.1016/j.cedpsych.2003.09.002

Bahr, P. R. (2010). Making sense of disparities in mathematics remediation: What is the role of student retention? *Journal of College Student Retention: Research, Theory & Practice*, *12*(1), 25–49. doi:10.2190/CS.12.1.c

Bartley, S. J., & Golek, J. H. (2004). Evaluating the cost effectiveness of online and face-to-face instruction. *Educational Technology & Society*, *7*(4), 167–175.

Benbunan-Fich, R., & Hiltz, S. R. (2003). Mediators of the effectiveness of online courses. *IEEE Transactions on Professional Communication*, *46*(4), 298–312. doi:10.1109/TPC.2003.819639

Bonham, B. S., & Boylan, H. R. (2011). Developmental mathematics: Challenges, promising practices, and recent initiatives. *Journal of Developmental Education*, *34*(3), 2.

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. doi:10.1002/pam.21728

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*(7), 775–801. doi:10.1007/s11162-007-9053-8

Clarke, T., & Hermens, A. (2001). Corporate developments and strategic alliances in e-learning. *Education+ Training*, *43*(4/5), 256–267. doi:10.1108/00400910110399328

Clinefelter, D. L., & Aslanian, C. B. (2015). *Online college students 2015: Comprehensive data on demands and preferences*. Louisville, KY: The Learning House, Inc.

Corbeil, J. R. (2003). *Online technologies, self-efficacy, self-directed learning readiness, and locus of control of learners in a graduate-level web-based distance education program*. (Unpublished doctoral dissertation). University of Houston, Houston, TX.

Delen, E., Liew, J., & Willson, V. (2014). Effects of interactivity and instructional scaffolding on learning: Self-regulation in online video-based environments. *Computers & Education*, *78*, 312–320. doi:10.1016/j.compedu.2014.06.018

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*. National Bureau of Economic Research.

Fullan, M. (2014). *Leading in a culture of change personal action guide and workbook*. San Francisco, CA: John Wiley & Sons.

Guglielmino, L. M., & Guglielmino, P. J. (2002). Learner characteristics affecting success in electronic distance learning. In H. B. Long Associates (Ed.). *Twenty-first century advances in self-directed leaning* (pp. 125–148). Boynton Beach, FL: Motorola University Press.

Hannafin, M. J., & Land, S. M. (1997). The foundations and assumptions of technology enhanced student-centered learning environments. *Instructional Science*, *25*, 167–202. doi:10.1023/A:1002997414652

Joyce, T., Crockett, S., Jaeger, D. A., Altindag, O., & O'Connell, S. D. (2015). Does classroom time matter? *Economics of Education Review*, *46*, 64–77. doi:10.1016/j.econedurev.2015.02.007

Kearsley, G. (2002). Is online learning for everybody? *Educational Technology*, *42*(1), 41–44.

McEwan, P. J. (2012). Cost-effectiveness analysis of education and health interventions in developing countries. *Journal of Development Effectiveness*, *4*(2), 189–213. doi:10.1080/19439342.2011.649044

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Retrieved from https://www2.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.pdf

Merisotis, J. P., & Phipps, R. A. (2000). Developmental education in colleges and universities: What's really going on? *The Review of Higher Education*, *24*(1), 67–85. doi:10.1353/rhe.2000.0023

Moore, M. (1987). Distance learning in the United States: The near future. *Distance Education*, *8*(1), 38–46. doi:10.1080/0158791870080103

Parsad, B., Lewis, L., & Greene, B. (2003). *Developmental education at degree-granting postsecondary institutions in fall 2000* (NCES 2004–010). US Department of Education. National Center for Education Statistics. Washington, DC: US Government Printing Office.

Pazzaglia, A. M., Clements, M., Lavigne, H. J., & Stafford, E. T. (2016). *An analysis of student engagement patterns and online course outcomes in Wisconsin*. REL 2016-147. Regional Educational Laboratory Midwest.

Proffitt, S. (2014). *Commercially Available or Home-grown: A Cost-effectiveness Analysis of K-12 Online Courses* (Doctoral dissertation). Retrieved from Virginia Commonwealth University Scholars Compass.

Rosenbaum, P. R. (2005). Sensitivity analysis in observational studies. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol. 4, pp. 1451–1461). Chichester, West Sussex: Wiley.

Sparks, D., & Malkus, N. (2013). *First-year undergraduate developmental coursetaking: 1999–2000, 2003–04, 2007–08*. Statistics in Brief. NCES 2013-013. National Center for Education Statistics.

Strong American Schools. (2008). *Diploma to nowhere*. Retrieved from http://www.broadeducation.org/

Wang, C. H., Shannon, D. M., & Ross, M. E. (2013). Students' characteristics, self-regulated learning, technology self-efficacy, and course outcomes in online learning. *Distance Education*, *34*(3), 302–323. doi:10.1080/01587919.2013.835779

Williams, M. (1996). Learner control and instructional technologies. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 957–983). New York, NY: Scholastic.

Xu, D., & Jaggars, S. S. (2014). Performance gaps between online and face-to-face courses: Differences across types of students and academic subject areas. *The Journal of Higher Education*, *85*(5), 633–659. doi:10.1353/jhe.2014.0028

Yen, H. J., & Liu, S. (2009). Learner autonomy as a predictor of course success and final grades in community college online courses. *Journal of Educational Computing Research*, *41*(3), 347–367. doi:10.2190/EC.41.3.e