
Routledge
Taylor & Francis Group

The Journal of Higher Education

ISSN: 0022-1546 (Print) 1538-4640 (Online) Journal homepage: http://www.tandfonline.com/loi/uhej20

# Interactive Online Learning on Campus: Comparing Students' Outcomes in Hybrid and Traditional Courses in the University System of Maryland

Matthew M. Chingos, Rebecca J. Griffiths, Christine Mulhern & Richard R. Spies

THE JOURNAL OF HIGHER EDUCATION
2017, VOL. 88, NO. 2, 210–233
http://dx.doi.org/10.1080/00221546.2016.1244409

Routledge
Taylor & Francis Group

# Interactive Online Learning on Campus: Comparing Students' Outcomes in Hybrid and Traditional Courses in the University System of Maryland

Matthew M. Chingos[a], Rebecca J. Griffiths[b], Christine Mulhern[c], and Richard R. Spies[d]

[a]Income and Benefits Policy Center, Urban Institute, Washington, DC, USA; [b]Center for Technology in Learning, SRI International, Arlington, VA, USA; [c]John F. Kennedy School of Government, Harvard University, Cambridge, Massachusetts, USA; [d]Educational Transformation, Ithaka S+R, New York, New York, USA

**ABSTRACT**
Massively open online courses (MOOCs) have received a great deal of attention, but little research exists on how they might fit into the existing system of higher education. We studied the impacts on learning outcomes of hybrid courses redesigned using online materials from MOOCs created on the Coursera platform and digital materials created by the Open Learning Initiative (OLI), relative to existing versions of the same courses. We found that student performance was about the same in both sections, as measured by pass rates and scores on common assessments. This finding held across a variety of disciplines and subgroups of students. We found no evidence supporting the worry that disadvantaged or academically underprepared students were harmed by taking hybrid courses with reduced class time. Despite the similar student outcomes produced by the two course formats, students in the hybrid sections reported considerably lower satisfaction with their experience.

**ARTICLE HISTORY**
Received 6 March 2015
Accepted 22 June 2016

**KEYWORDS**
Hybrid courses; MOOCs; online learning



## Introduction

Massively open online courses (MOOCs) made a big splash when they first launched in the United States in 2011, with providers such as Coursera, edX, and Udacity generating enthusiasm that students everywhere could take high-quality online courses for free, many from professors at the nation's most elite universities. *The New York Times* called 2012 "The Year of the MOOC," with more than 1.7 million students enrolling in courses through Coursera. Anant Argawal, a founder of edX, predicted that "students will one day arrive on campus with MOOC credits the way they now do with advanced placement credits" (Pappano, 2012). In his book *Higher Education in the Digital Age*, Bowen (2013) wrote that "it seems clear that MOOCs have an extraordinary capacity

**CONTACT** Christine Mulhern mulhern@g.harvard.edu
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uhej.

to improve access to educational materials from renowned instructors in various subjects for learners throughout the world" (pp. 60–61). Proponents have argued that MOOCs could better educate students than many existing courses and some have predicted they would disrupt the centuries-old model of higher education in the United States (Mazoue, 2013).

The hype around MOOCs has since calmed, as early hopes and fears that MOOCs would provide a viable alternative to traditional education have not borne out (Borden, 2014). Existing evidence has suggested that MOOCs have thus far served largely as a source of free adult education for individuals who are already quite educated (Christensen et al., 2013). Still, they remain popular: More than 17 million students have enrolled in 1,756 Coursera courses from 147 partner institutions (Coursera, 2016; Shah, 2015).

MOOC providers set out to produce a set of high-quality content that could be delivered exclusively over the Internet to a large group of students. MOOCs typically consist of lecture videos, readings, quizzes, and papers or project assignments. There are no strict guidelines about what must be included in these courses, and they vary in length and the amount of content covered. By definition, they are open to anyone, with no enrollment cap, and they can be fully accessed and completed online (Educause Library, 2015). Coursera, the largest MOOC platform, sought top professors at elite universities to teach their courses as MOOCs; edX was founded by Harvard and Massachusetts Institute of Technology, with a focus on hosting MOOCs based on their own courses (edX, 2015; Rivard, 2013). Thus, the initial MOOCs were mostly produced by highly regarded professors and academic institutions, which led many to believe that MOOCs would be of superior quality to other online content and courses (Fain, 2012; Stokes, 2013). However, little work has actually been done to assess the quality of MOOCs, and most attributions of quality stem from people focusing on the institutions and professors.

Developing high-quality online content is not unique to MOOCs—other providers had been developing content for online and hybrid courses (which blend online and in-person instruction) for years before MOOCs came on the scene.[1] One goal of these content providers has been to improve student learning and progress toward earning a postsecondary degree (Fain, 2012; Open Learning Initiative, 2015; Stokes, 2013). In light of recent research showing increasing time to a degree and rising costs for a college degree, it is important to find ways to improve student learning and postsecondary degree attainment (Bound, Lovenheim, & Turner, 2012; Mulhern, Spies, Staiger, & Wu, 2015; The White House, 2015).

Recent literature on flipped classrooms has suggested that courses that require students to study static material outside of class and then participate in hands-on activities in class may lead to better student engagement and learning than traditional courses (Bishop & Verleger, 2013). In addition, some researchers have encouraged colleges to adopt hybrid courses to take advantage of the increased access that online environments can provide while maintaining the face-to-face contact with the instructor and classmates that many deem important to student success (Brown, 2001; Oblender, 2002). Furthermore, these courses may reduce costs because they require the instructor to be in the classroom for less time and lecture videos can be used multiple times with potentially an unlimited number of students (Bowen, 2013).

Thus, a natural question to ask is whether MOOC content can be repurposed for use in hybrid courses on traditional college campuses in ways that enhance learning outcomes. MOOCs have the potential to expand access to elements of the elite education that only a fraction of students are able to receive. Given the high regard for the teaching and learning that occur at these elite institutions, it is plausible that extending these learning experiences to other students could improve their learning outcomes.

We set out to answer this question with public university campuses in the University System of Maryland (USM).[2] The primary research question we addressed was how instructor use of online materials to redesign traditional courses impacts student learning outcomes. The online materials used in our study included four different courses provided by Coursera and three instances of the biology course designed by Carnegie Mellon University's OLI, a creator of interactive online courses.

Our study is the first of its kind to systematically examine the impact of redesigned hybrid courses using materials from multiple online content creators across a range of disciplines on public university campuses. There is little high-quality research concerning the impact of online education materials on student learning in postsecondary education. A large number of studies exist, but very few make careful comparisons of students taking the same courses in different formats using either random assignment or quasi-experimental methods to adjust for differences in baseline characteristics (Means, Toyama, Murphy, Bakia, & Jones, 2010). In addition, it is important to separate the impacts of hybrid courses and online courses, because one may expect even a small amount of face-to-face time to have substantial impacts on student learning (Means et al., 2010).

The studies most relevant to the present study have revealed no significant difference in student learning outcomes in hybrid and traditional courses. The study by Bowen, Chingos, Lack, and Nygren (2014) is the only study of hybrid courses that randomly assigned (volunteer) students to ensure that the treatment and control groups were balanced (in expectation) on pretreatment characteristics. Bowen et al. (2014) studied seven instances of the OLI statistics course on six different public university campuses. They found that outcomes, measured by pass rates and assessment scores, were largely similar across the two formats and that students in the hybrid format spent 18% less time learning the material. However, students reported lower satisfaction with the hybrid course than with the traditional course.

More positive results were found in a meta-analysis of quasi-experimental studies by Means et al. (2010). This analysis indicated that students who took a hybrid class, on average, performed significantly better than students who took a traditional face-to-face course. However, according to Lack (2013), "very few [of the studies] have direct relevance to the large public universities or the broad-based community colleges that educate such a high fraction of the country's student population" (p. 4). Only five of the studies contained more than 400 students, and many focused on graduate students or medical fields.

Studies that allowed students to select into a format but controlled for student characteristics have shown mixed results. The different results may be due to limitations in the study designs or the course subjects, because most studies only focus on one subject.

Two studies with small samples and different subjects indicated opposite results. Chenoweth, Ushida, and Murday (2006) found that students in hybrid language courses performed at least as well as students in the traditional versions and better in two cases. Conversely, Verhoeven and Rudchenko (2013) reported lower test scores for students in a hybrid version of an introductory microeconomics course compared with students in the face-to-face version.

Two studies using student grades as an outcome showed that students in the hybrid version earned lower grades than their peers in the face-to-face version, but grades are often considered a more subjective measure than common assessments (Burns, Duncan, Sweeney, North, & Ellegood, 2013; Tanyel & Griffin, 2014). Xu and Jaggars (2011, 2013) also used transcript data but found that students at community and technical colleges were equally as likely to complete hybrid courses as they were to complete face-to-face courses, while controlling for student characteristics. Unlike the other studies, they looked at a variety of subjects, but their data were not fine enough to identify the subject areas in which hybrid courses are offered.

In attempt to remove the issue of student and teacher selection bias, Kwak, Menezes, and Sherwood (2014) replaced 2 weeks in the middle of a traditional face-to-face statistics class with a blended format. Student quiz scores were not significantly different during the 2 weeks with blended instruction, as compared with the rest of the course. However, 2 weeks is a short period on which to base these results, and quiz scores were the only outcome measured. It is possible that what students learned or how they would perform on more substantial assessments is not reflected in these quiz scores. In addition, instructors in our study mentioned that it took students a couple weeks to adjust to the hybrid format, so student performance may increase after the first few weeks of this format.

While the present study does not explore fully online courses, the materials used in online courses may shed light on the potential for these resources to impact student learning. Figlio, Rush, and Yin (2013) found that students in the online-only section of an undergraduate microeconomics course had learning outcomes that were not statistically distinguishable from the live-lecture group, but the online-only treatment had negative impacts for Hispanics, men, and lower-achieving students. Xu and Jaggars (2014) found a larger gap in student performance between online and face-to-face courses for students who are male, younger, Black, or have lower levels of prior academic performance. Given these findings, we looked at differences in student outcomes by subgroups of students but did not find patterns similar to these studies.

There is also some evidence that reducing the time students spend in class does not substantially harm their learning outcomes. Joyce, Crockett, Jaeger, Altindag, and O'Connell (2014) found that students randomly assigned to an introductory microeconomics section that met twice per week instead of once per week performed slightly better on common assessments, but the difference was small. Thus, while some face-to-face time may be important in increasing student learning, it is possible that more time does not always lead to better outcomes. Therefore, hybrid courses may not harm student learning.

The present literature has yet to settle the debate on whether or not hybrid learning is beneficial to student learning outcomes. The growing prevalence of courses that incorporate technology makes it important to better understand how hybrid courses impact student learning and success (Allen & Seaman, 2014). Furthermore, the broad range of online content used in hybrid courses increases the importance of understanding how content from MOOCs and OLI fare in these courses.

The present study extends the literature on the impact of hybrid learning models in higher education in at least two significant ways. First, our study is the only systematic empirical evaluation of the use of materials from MOOCs (and other content creators) on traditional campuses.[3] Second, our study is the first to use common assessments of learning outcomes to examine the impact of hybrid learning models across a range of academic disciplines, including biology, communications, computer science, precalculus, and statistics, and with multiple online platforms. Several of the studies previously discussed only examined a course in a single discipline, and collectively, they covered only four subjects. The studies that used observational transcript data did not assess the subjects in which hybrid courses were more likely to appear and thus did not test whether outcomes differed within a specific subject between hybrid and traditional courses.

We found that students in the hybrid sections did as well as students in the traditional sections in terms of pass rates and learning assessments—a finding that held across all disciplines and subgroups of students. We found no evidence supporting the worry that disadvantaged or academically underprepared students were harmed by taking hybrid courses. Despite the similar student outcomes produced by the two course formats, students in the hybrid sections reported considerably lower satisfaction with their experience.

## Course selection and data

This research aimed to understand how hybrid courses incorporating existing online content from Coursera's MOOCs or Carnegie Mellon's OLI compared to traditional face-to-face courses. We worked with instructors at three institutions in the USM to set up side-by-side comparisons of hybrid and traditional versions of seven different courses in the fall 2013 semester.[4] Four of the courses, in communications, computer science, precalculus, and statistics, incorporated one of Coursera's MOOCs, and three courses used ordinary least square regressions biology course. The three participating institutions include an Historically Black College/University located outside of Washington, DC, a large public institution located in the northern suburbs of Baltimore, and a small university located on Maryland's Eastern Shore.

The particular instructors and subjects were chosen during the winter prior to the fall semester when the course was taught. Staff from USM and the research team presented the project at each USM campus to faculty and administrators. All instructors of undergraduate courses were invited to submit a proposal to participate in the study.[5] Institutions and instructors were offered money to cover expenses related to the project and time spent on it, assistance gaining permission to use restricted MOOC materials, and technical support for their courses.[6]

Proposals needed to include (a) a clear vision for how the use of online materials could address a pedagogical objective or solve a problem for the course or program, (b) a commitment to make the online materials required for a substantial portion of the course and to use at least 40% of the materials available, (c) a description of the desired online materials, (d) a description of the course, (e) expected enrollment, and (f) an indication of instructors' willingness to administer a common posttest that would count toward students' grades in all control and treatment sections and willingness to participate in other data collection activities.

The researchers selected the final group of courses based on the proposals that met the criteria to be included in a side-by-side comparison. These courses needed to (a) be sufficiently large, (b) have comparable control sections, (c) have online materials that fit well with the course, (d) be granted permission to use the online materials (which was gained from Coursera and professors), and (e) present proposals that fit within the required terms.[7] Thirty instructors submitted proposals, though several did not meet all of the

**Table 1.** Description of courses.

| Course | Tech. | Sections | | Students | | Minutes/week | | Total students |
|---|---|---|---|---|---|---|---|---|
| | | T | C | T | C | T | C | |
| Biology A | OLI | 6 | 1 | 146 | 147 | 50 | 150 | 293 |
| Biology B | OLI | 3 | 2 | 88 | 80 | 50 | 150 | 168 |
| Biology C | OLI | 2 | 2 | 239 | 239 | 50 | 50 | 478 |
| Communications | MOOC | 4 | 7 | 104 | 103 | 80 | 160 | 207 |
| Computer science | MOOC | 4 | 5 | 93 | 84 | 75–100 | 150 | 177 |
| Precalculus | MOOC | 2 | 3 | 55 | 67 | 120 | 220 | 122 |
| Statistics | MOOC | 2 | 3 | 64 | 95 | 150 | 150 | 159 |

*Note*. OLI = Open Learning Initiative; MOOC = massively open online course. "T" refers to the treatment group and "C" refers to the control group. "Minutes/week" refers to the length of in-person course meetings.



criteria. As a result, seven introductory courses with relatively large enrollments were selected.[8] Coursera worked with the researchers and instructors to select MOOCs that were deemed as being high quality because they had been run at least once before, had been well received by public audiences, and did not run into major pedagogical or technical problems.

Table 1 lists, for each course, the number of treatment (hybrid) and control (traditional) sections, the number of students enrolled, and the amount of class time each week. Five of the seven courses had treatment sections that met for less time than the traditional sections. In these cases, the online materials replaced some of the time and materials that the instructors taught in the traditional sections. In the other two courses, meetings times were the same and online materials were used more as a supplement rather than a replacement for class time. Students had access to the MOOC and OLI content and were assigned to complete it as a part of their homework. [9]

MOOCs contained video lectures as the primary source of content, and most contained quiz questions and readings. The OLI biology course contained text-based explanations and interactive practice problems. The instructors were free to choose how to incorporate the online content into their courses. All four courses using MOOCs used the videos and some of the quiz questions. The instructors typically assigned students to watch a set of videos (or review the OLI content) and complete the corresponding quiz before attending class. Students were expected to learn most of the content outside of class, and quizzes were used to incentivize students.[10] Class time was used to discuss the content learned at home and to engage in practice problems and hands-on activities.

The control sections covered the same content as the treatment sections but relied on different materials or the instructor to provide the content.[11] In a few cases, the courses differed from the versions taught in previous years because instructors or departments wanted to redesign the curriculum or cover new content. Some of the control section instructors spent considerable time redesigning their courses to cover the content that the other students would learn with the MOOC/OLI. Because posttests were part of students' grades, instructors were careful to ensure that students in all sections were treated fairly in terms of preparation for these tests.[12]

The research team gathered background data, including SAT scores, cumulative grade point average (GPA) at the beginning of the semester, race/ethnicity, gender, parental income (for aid applicants), and date of birth, from administrative records maintained by the institution.[13] For each course, the instructors of the hybrid and traditional sections had to agree on a common form of assessment. Most offered both a common final exam across the sections as well as a pretest and posttest.[14] Instructors provided the results of these assessments to the research team, and course grades were gathered from administrative records.

Additional information was gathered from surveys administered to students at the beginning and end of the semester. The initial survey gathered background information not available in administrative records, such as parents' education and students' educational aspirations. The final survey asked students about their experiences in the course.[15] We made only limited use of the survey data given that response rates were 77% on the initial survey and 75% on the final survey.[16] Specifically, we used self-reported data on family income to supplement administrative data on aid applicants and student course evaluations to examine their opinions of the course.

## Research design

The ideal research design for estimating the causal effect of the hybrid format would be to randomly assign students and instructors to one of the two formats and then compare average outcomes across formats. Random assignment was not possible for logistical reasons. Randomly assigning instructors was not possible because in most cases, the hybrid courses were designed by volunteer instructors. Randomly assigning students would have been logistically challenging (e.g., hybrid- and traditional-format sections would need to be offered at the same time so a student could attend either) and would have sharply reduced our sample sizes to the subset of students willing and able to be randomly assigned.

Students registered for the seven courses in our study as they normally would. Selection bias would have been an issue if students chose their courses based on the instructor, the time the class met, or the format listed in the course catalogue. However, there are a few reasons to believe that selection bias was not substantial in this context. Four of the courses did not indicate on the course schedule that some sections were incorporating a MOOC or OLI, so in these cases, students were unaware of the section in which they were enrolling. In two of the other courses, students were freshmen and had little say in the sections in which they enrolled because of limited availability and advisor placements.[17] Consequently, differences in student characteristics between the two formats might not have been as substantial as would they would have been the case if they all were able to purposefully choose their preferred format.

This hypothesis is largely supported by the data. Table 2 reports summary statistics for the treatment and control sections. We also report within-course differences (and corresponding *p* values) between the treatment and control groups, averaged across all of the courses. These differences were calculated by regressing each variable listed in Table 2 on a treatment indicator and dummy variables identifying each of the courses.

These data show that the 1,598 students in our study are a diverse group, a fact that holds true in both formats of the courses. Students' SAT performance, a measure of their academic preparation, was nearly identical, on average, in the two groups.[18] Both groups were roughly half White, one third Black, and 60% female. The average characteristics of the treatment and control groups were not statistically significantly different in most cases, although we can reject a null hypothesis of no correlation between all of these characteristics and section format.

These data suggest that students' decisions about whether to take a course in a hybrid format bore little relation to their academic preparation and personal backgrounds. It is possible that these decisions hinged more on

**Table 2.** Summary statistics, students.

| | Traditional ($N$ = 820) | Hybrid ($N$ = 778) | Adjusted difference | $p$ value corresponding to adjusted difference |
|---|---|---|---|---|
| SAT Math | 511 | 514 | 4.11 | .11 |
| SAT Verbal | 511 | 510 | −0.45 | .92 |
| SAT missing | 21% | 21% | 0.00 | .96 |
| Cum GPA | 2.82 | 2.85 | 0.01 | .85 |
| Cum GPA missing | 51% | 55% | 0.05 | .47 |
| Race/ethnicity | | | | |
| White | 50% | 51% | 0.01 | .39 |
| Black | 31% | 34% | 0.01 | .57 |
| Hispanic | 4% | 4% | 0.01 | .39 |
| Asian | 5% | 4% | −0.01 | .29 |
| Other/missing | 9% | 7% | −0.02 | .12 |
| Female | 61% | 60% | −0.02 | .38 |
| Parents' income | | | | |
| Less than $50,000 | 15% | 17% | 0.03 | .11 |
| $50,000–$100,000 | 20% | 21% | 0.02 | .10 |
| More than $100,000 | 37% | 33% | −0.03 | .13 |
| Missing | 28% | 29% | −0.01 | .47 |
| Age | 20.0 | 19.8 | −0.22 | .23 |
| Pretest score | 47% | 52% | 0.04 | .02 |
| Pretest missing | 33% | 27% | −0.07 | .12 |

*Note*. GPA = grade point average. The reported $p$ values of the traditional–hybrid differences control for course dummies and are based on robust standard errors adjusted for clustering by section. A regression of treatment group on all of the variables listed here rejects the null of 0 coefficients for all variables with $p$ = .00 (excluding pretest scores, $p$ = .02).

factors such as scheduling convenience than on the format of the course. It is also worth noting that two of the courses (biology C and statistics) had equal amounts of face-to-face time in the traditional and hybrid sections.

One limitation of our data is that they do not include students who switched sections or dropped the course within the first couple of weeks of the semester. For each course, the instructors estimated that no more than 5 to 10 students switched sections because they were limited by available seats (most sections were full). Given the small number of students involved, we think it is unlikely that our results are substantially biased by our inability to identify students who dropped the course or switched sections early in the semester. However, the results could be biased if students who were uncomfortable with the hybrid version switched into the treatment section or if students in the traditional section thought the hybrid version may be easier because it met for less time.[19] Anecdotally, instructors did not think this was a big problem.

Table 3 shows descriptive statistics about the sections and the instructors who led them. The typical section enrolled just more than 75 students in both formats, although the traditional sections met for about an hour more each week than did the hybrid sections. About half the instructors in both groups were tenure-track faculty, and nearly all were employed full-time by their institution. The two groups of instructors had similar amounts of teaching experience in terms of total years, but the traditional instructors had taught the course more times and also had more experience at their current institutions.[20] Despite the similarity of measured instructor characteristics, it is of course possible that instructors chose the format in which they would be more effective. Consequently, our results should be interpreted as the effect of hybrid instruction as implemented by the volunteer instructors, not as the effect that would be obtained if all instructors were forced to teach in this format.

We estimated the effect of the hybrid format by comparing outcomes of students in the control and treatment groups. Specifically, we estimated regressions of the form:

$$Y_i = \beta_0 + \beta_1 * Hybrid_i + \alpha * X_i + \varepsilon_i,$$

**Table 3.** Summary statistics, sections/instructors.

| | Traditional | Hybrid |
|---|---|---|
| Weekly face-to-face minutes | 126 | 72 |
| Section size | 76 | 77 |
| Instructor characteristics | | |
| Tenure track | 52% | 54% |
| Full-time | 91% | 100% |
| Taught with tech before | 81% | 55% |
| Years of teaching experience | 14.5 | 13.5 |
| Total times taught course | 18.7 | 13.4 |
| Times taught course here | 18.2 | 7.7 |

*Note*. Statistics are weighted by student enrollment in each section.

where $Y_i$ is the outcome of student $i$, $\beta_0$ is a constant, *Hybrid* is a dummy variable identifying whether the student was in a hybrid section of the course (with coefficient $\beta_1$), $X_i$ is a vector of control variables with coefficient vector $\alpha$, and $\varepsilon_i$ is the error term that we adjusted by clustering by section. Control variables included SAT math and verbal scores, cumulative GPA at the beginning of the semester, race/ethnicity, gender, parental income, and age. Missing values were identified by dummies for each variable, with continuous variables imputed using an arbitrary value (in this case, 0).

We estimated the equations using OLS.[21] We focused on three outcome variables: whether the student passed the course, the score on the posttest or final (only using low-stakes posttests for courses that did not administer a common final exam), and the grade in the course (measured on a 4-point scale).[22] We focused on pass rates because they determine whether students earn credit for the course and on the assessment outcomes because they are least subject to bias from differential grading standards across instructors.

## Results

### Main results

We found that students in the hybrid sections performed the same as their peers in the traditional sections. Table 4 shows that the pass rates of hybrid-format students were about 4 percentage points higher than those of traditional-format students, 83% of whom passed the course ($p$ = .281, without controls). A similar pattern of results prevailed on common assessments administered at the end of the course—the final exam or a subset of common exam questions. Some courses also administered a posttest, which we used as an alternate measure in our analyses.[23] The hybrid-format advantage in exam scores was 2 percentage points on a traditional-format mean of 70% correct ($p$ = .097).[24] In the case of the final assessment scores (but not pass rates), the difference between traditional and hybrid sections was statistically significant

**Table 4.** Difference between outcomes in hybrid and traditional sections, all courses.

| | Pass | | Posttest/final | | Course grade | |
|---|---|---|---|---|---|---|
| Hybrid estimate | 0.036 | 0.030 | 0.024* | 0.024* | −0.002 | −0.024 |
| | (0.033) | (0.032) | (0.014) | (0.012) | (0.176) | (0.169) |
| Controls? | No | Yes | No | Yes | No | Yes |
| Observations | 1,594 | 1,594 | 1,187 | 1,187 | 1,564 | 1,564 |
| Control mean | 0.83 | 0.83 | 0.70 | 0.70 | 2.21 | 2.21 |
| Control *SD* | | | 0.16 | 0.16 | 1.34 | 1.34 |

*Note*. * $p < 0.1$. Robust standard errors adjusted for clustering by section appear in parentheses. All models include dummies for each course. Controls include SAT math and verbal scores, cumulative grade point average at the beginning of the semester, race/ethnicity, gender, parents' income, and age. Missing values are identified by dummies for each variable, with continuous variables imputed using an arbitrary value (in this case, 0). The course grade is measured on a 4-point scale.

at the 10% level. We found little differences in course grades, a more subjective measure on which we report but do not focus ($p$ = .989).

In the absence of an experiment in which students are randomly assigned to one format or the other, we cannot be completely sure that these differences in student outcomes indicate the true effect of the different section formats and not some other factors such as different types of students choosing one format over the other. However, there are two key pieces of evidence that increase our confidence in the results. First, as discussed, the characteristics of students captured in our data were broadly similar between the two formats. Second, when we accounted for these factors in the analysis, our estimate of the hybrid effect did not change substantially. Although this fact is far from dispositive, it provides some confidence that student characteristics are not substantially biasing the results. Table 4 shows that controlling for SAT scores, cumulative GPA at the beginning of the semester, race/ethnicity, gender, parental income, and age changes our estimate of the effect on pass rates from 3.6 to 3.0 percentage points ($p$ = .355), leaving the estimate for the posttest/final unchanged at 2.4 percentage points ($p$ = .064).

It is certainly possible that traditional- and hybrid-format students differed in their unmeasured traits, such as motivation or perseverance, which were not captured by the characteristics we observed. However, the similarity of the two groups at baseline and the fact that including control variables in the analysis made little difference gives us some confidence in the results in the absence of random assignment (or some other source of random variation in section selection).

### *Robustness checks*

We conducted several additional analyses to check on the robustness of our main results. These estimates are reported in Table A1. The first row displays our main results (with controls included). The second row adds controls for section size and two instructor characteristics: tenure-track status and years of teaching experience. This analysis is a crude attempt to adjust for any differences in the kinds of instructors that volunteered to teach a hybrid section as part of our study, relative to the instructors who continued to teach in the traditional manner. The estimated effects only changed modestly, and they indicated potentially larger effects than our main results.

The third row shows results that add a control for students' scores on a pretest administered at the beginning of the semester. Pretest scores were modestly higher in the hybrid sections, where students scored an average of 52% correct, compared with the traditional sections, where the average score was 47% (Table 2). However, 27% of hybrid-format students did not take the pretest, as compared with 33% of traditional-format students. (Only five of the seven courses administered a pretest.) We did not use the pretest as a control in our preferred results due to the difference in missing data rates between the groups. However, the estimated hybrid effect was only slightly smaller when we included this variable as a control.

Table A2 shows the results (with controls) by course as well as features of each course. When we averaged the results for the four courses in which students did not know the format for which they were registering, the point estimates were all very close to 0, but the results were not statistically distinguishable from our main results given the standard errors. We also examined the results separately for Coursera and OLI and found somewhat larger positive effects for OLI and null effects for Coursera, although the two effects were not statistically distinguishable from each other. Finally, excluding the two courses that did not reduce face-to-face time in the hybrid sections (statistics and biology C) produced results that were slightly larger than our main results. Thus, the slight differences across courses did not appear to be key drivers of our overall finding of no significant differences in student outcomes.

## Effect heterogeneity

Our main results combined data from all seven courses, which varied significantly in terms of subject matter, design of the hybrid format, and student and campus characteristics. Are these average effects a combination of some large, positive effects and other large, negative effects? Our data suggest that the results are fairly consistent across courses (Table A2).

In general, the results for pass rates and final assessment scores indicate a mix of small positive and small negative results. In most cases, these results were not statistically distinguishable from a 0 effect, which is not surprising given that each individual test was relatively small. However, there are a few exceptions. There was a large difference in pass rates (20 percentage points) favoring the hybrid sections of the biology A course. (If this course is excluded from the average result, the estimated effects are close to 0.) Discussions with campus faculty indicated that this finding was due to very stringent grading standards imposed by the instructor of the traditional section, which have historically led to low pass rates. In this case, we cannot tell if the results were driven by an instructor effect or a format/software effect. However, the hybrid-format students also performed better on the common exam questions, a measure that should not be affected by instructor grading standards. There was also a significant positive estimated effect on final exam scores in the precalculus course of about 10 percentage points.

There were a number of smaller estimated hybrid effects, both negative and positive, but given the precision of the analysis for an individual course, we cannot be confident about the direction much less the size of the effect.

The course-specific results took into account a number of student characteristics, as noted earlier. However, we obtained qualitatively similar results when we looked just at raw differences in outcomes.[25] In other words, there did not appear to be a substantial amount of bias in the results due to different kinds of students choosing the hybrid versus traditional sections, even when we looked separately at each course rather than averaging them all together.

We next estimated hybrid effects, averaged across all seven courses, for various subgroups of students defined in terms of demographics and academic preparation. This set of analyses tested whether certain groups of students fare less well than other groups with the hybrid format. These results, in Table 5, clearly show that our main finding of null or weakly positive effects of the hybrid format holds across a range of student groups. Table 5 shows a clear pattern of results that range from close to 0 to modestly positive. Of the 26 estimates in the table, only 1 is negative (and it is very small: −0.003). Perhaps, most significantly, we do not find any evidence that poorly prepared students, as identified by below-average SAT scores, were harmed by the hybrid format.

The worry that disadvantaged students are most likely to be harmed by technology replacing some face-to-face time was also not borne out by our data. If anything, we found the opposite. For example, the point estimates

**Table 5.** Difference between outcomes in hybrid and traditional sections, by student subgroup.

| | Pass | Posttest/final | | Pass | Posttest/final |
|---|---|---|---|---|---|
| White/Asian | 0.037 | 0.031** | Parents' income | 0.099* | 0.060** |
| | (0.029) | (0.012) | less than $50,000 | (0.050) | (0.025) |
| | 882 | 677 | | 258 | 204 |
| Black/Hispanic | −0.003 | 0.014 | Parents' income | 0.037 | 0.040*** |
| | (0.048) | (0.024) | $50,000–$100,000 | (0.045) | (0.010) |
| | 582 | 417 | | 327 | 254 |
| Female | 0.039 | 0.032** | Parents' income | 0.022 | 0.016 |
| | (0.033) | (0.014) | more than $100,000 | (0.024) | (0.015) |
| | 964 | 712 | | 556 | 417 |
| Male | 0.018 | 0.010 | Neither parent | 0.094 | 0.038 |
| | (0.036) | (0.017) | has bachelor's | (0.074) | (0.024) |
| | 630 | 475 | | 348 | 254 |
| First-year student | 0.052 | 0.027* | At least one parent | 0.014 | 0.028** |
| | (0.032) | (0.014) | has bachelor's | (0.023) | (0.012) |
| | 628 | 514 | | 657 | 499 |
| Not a 1st-year | 0.030 | 0.031 | SAT less than 1000 | 0.019 | 0.025 |
| student | (0.041) | (0.019) | | (0.039) | (0.018) |
| | 966 | 673 | | 852 | 625 |
| | | | SAT 1000 or higher | 0.040 | 0.028** |
| | | | | (0.032) | (0.012) |
| | | | | 742 | 562 |

*Note*. * $p < .1$. ** $p < .05$. *** $p < .01$. Robust standard errors adjusted for clustering by section appear in parentheses. Numbers of observations appear in italics. All models include controls for SAT math and verbal scores, cumulative grade point average at the beginning of the semester, race/ethnicity, gender, parents' income, and age. Missing values are identified by dummies for each variable, with continuous variables imputed using an arbitrary value (in this case, 0).

were largest for students from the lowest-income families and smallest for students from higher-income families, and the difference between the top and bottom income groups in estimated effects was statistically significant at the 5% level for posttest/final ($p$ = .046) but not for pass rates ($p$ = .115). It is important to interpret these results cautiously for the reasons stated earlier and because it is difficult to measure income accurately (much of our data were self-reported by the students on a survey).

### Student evaluations

Our analysis of administrative data on student outcomes—pass rates and scores on common assessments—showed that hybrid-format students did the same, on average, as their peers in the traditional sections. However, data from surveys we administered at the end of the courses suggest that talking to students would give the opposite impression. Students were asked several questions about the course, and we placed their responses on a 5-point scale, where 1 is the worst rating and 5 is the best. Table 6 shows average student responses on four of these measures in the hybrid and traditional sections.

Overall, traditional-format students gave the course a higher rating than did hybrid-format students and said they felt they learned more. But there were no significant differences between formats in students' reports on the difficulty of the course or how much it raised their interest in the subject matter.

We also examined whether these differences in student satisfaction varied across courses and student subgroups.[26] We found a substantial negative perception of the hybrid format (a difference of roughly 1 point on the 5-point scale) in four of the seven courses, with the other three showing only small differences. When we looked at perceptions of the same 13 student subgroups we examined earlier, we found a negative effect of the hybrid format on overall satisfaction for all of them, with little variation in the size of the difference.

Finally, we analyzed how much time students spent working on course material. Figure 1 shows that the hybrid sections met for roughly 1 hour less than did the traditional sections. But hybrid-format students reported

**Table 6.** Difference between student evaluations in hybrid and traditional sections, all courses.

| | Hours/week | Overall | Interest | Learn | Difficulty |
|---|---|---|---|---|---|
| Hybrid estimate | 0.2 | −0.6*** | −0.1 | −0.4*** | 0.1 |
| | (0.4) | (0.1) | (0.1) | (0.1) | (0.1) |
| Observations | 1,126 | 1,196 | 1,199 | 1,191 | 1,200 |
| Control mean | 4.3 | 3.5 | 2.9 | 3.6 | 3.2 |
| Control *SD* | 3.6 | 1.0 | 1.2 | 1.0 | 1.0 |

*Note*. *** $p < .01$. Robust standard errors adjusted for clustering by section appear in parentheses. All models include dummies for each course.

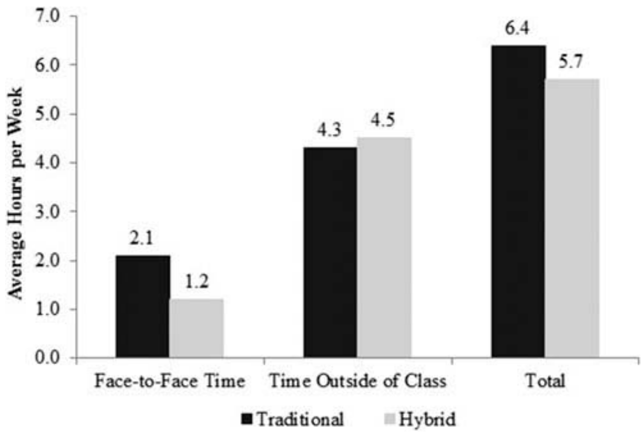


Figure 1. Student time spend on course.

spending about the same amount of time on coursework outside of class, on average, as the traditional students. Taken together, these findings suggest that students in the hybrid sections earned similar grades and pass rates, despite being less satisfied and spending 11% less time on the course.

## Limitations

Our results are subject to two important qualifications. First, we were not able to randomly assign students or instructors to section formats. It is possible that students chose the format in which they expected to be most successful or the format they thought would be easiest, although measured student characteristics in the hybrid and traditional sections were similar, on average. The lack of data on students who withdrew from the class in the first few weeks also limits our findings to the population of students who enrolled in a course through the drop deadline.

The results should be interpreted in the context of the volunteer instructors who designed the hybrid courses and should not be extrapolated to contexts where nonvolunteer instructors are assigned to design courses or teach in a hybrid format. The fact that we were not able to randomly assign volunteer instructors to the two formats leaves open the possibility of bias, but the results are not sensitive to controlling for observed instructor characteristics. Additionally, all but one of the instructors were teaching the newly designed version of the course for the first time, with the outcomes of their students compared to those of students taking a traditional version that had usually been taught several times before. For this reason, we might expect our results to understate the hybrid-format effects that would be obtained if the hybrid-format instructors had more experience with the online course materials and format.

Second, the intervention we evaluated was not a specific product, platform, or mode of instruction, but rather the average impact across seven courses of providing instructors with wide latitude to redesign traditional courses using technology. Instructors chose which MOOC or OLI course they wanted to use, and Coursera guided instructors toward the most "successful" MOOCs.[27] Instructors also had the freedom to design their own assessment tool, so long as it was the same for the treatment and control sections of that course. Whether we would obtain similar results if the same "intervention" were applied to seven different courses is an open question. The fact that we did not obtain wildly different results for individual courses suggests that our findings may hold in other contexts, but the imprecision of results for individual courses limits our ability to examine the extent of variation in results.

## Conclusion

Our findings add empirical weight to an emerging consensus that technology can be used to replace some classroom time without compromising student outcomes (Means et al., 2010). Students in the hybrid sections did as well as students in the traditional sections in terms of pass rates and assessment scores—a finding that held across disciplines and subgroups of students. We found no evidence supporting the worry that disadvantaged or academically underprepared students were harmed by taking hybrid-format courses. Previous studies reporting that disadvantaged groups perform worse primarily focused on online courses, which suggests that some face-to-face time is especially important for these students.

These findings are significant given that instructors were teaching the redesigned courses and using new technology platforms for the first time with, on average, just more than half as much class time as traditionally taught sections. Our findings are also broadly consistent across multiple subjects and universities. Thus, our results generalize to hybrid courses to a greater extent than do results from the existing more narrowly focused studies. Furthermore, the use of common assessments decreases the likelihood that the results were swayed by instructors' subjective evaluations. However, we note the limitations in common assessments, which varied in their use of multiple-choice and open-ended questions. The disagreement between the self-reported measures of student learning (which were lower in the hybrid sections) and the assessments might have been the result of the insensitivity of the assessments to some dimensions of student learning.

Continued experimentation with the design of hybrid courses will need to address many organizational and external issues to take full advantage of these technologies. Instructors need assurances that these materials will continue to be available for use, technically reliable, and adequately supported for multiple iterations. At present, however, there is uncertainty about the future direction of MOOC platforms, which have not developed institutional licensing models. Open educational resources often raise concerns about sustainability. Faculty must have the ability to customize course content, and this too is not ensured. They need access to a greater selection of online content and better indexing so that these materials are easier to locate, thereby reducing the amount of work required to fit the online materials with their courses and improving the student experience. More modular, flexible online resources would also make the implementation process easier.

The use of technology to redesign large introductory courses has the potential to reduce costs in the long run by reducing the time instructors spend planning and delivering courses. But not surprisingly, we found that redesigning courses to incorporate existing online content has significant start-up costs. The data we collected from instructors indicated that designing courses using MOOCs or OLI is a substantial undertaking and can take approximately 150 hour to 175 hour, with wide variations between instructors (Griffiths, Chingos, Mulhern, & Spies, 2014).

We do not have conclusive evidence of how use of these technologies on a large scale would impact costs, but offsetting these initial costs in the long run may well be possible over multiple iterations of efforts like those we have studied here. As participants gain experience using these tools and integrating them into college and university structures, it may be possible both to improve outcomes and reduce costs. Cost reductions may also be achieved by improving outcomes at the same cost per student, which would lead to lower costs per graduated student (Bailey, Jaggars, & Jenkins, 2015). Substantial cost savings would, we believe, require more strategic use of these tools across departments and institutions. Effective leadership would be essential —significant cost savings are not going to "just happen" on their own.

In addition, hybrid courses have the potential to improve access to higher education and reduce costs to students by providing them with more flexible schedules. Thus, even maintaining learning outcomes in specific courses may lead to better results overall and reduced costs if more students are able to take courses or earn a degree more quickly.

Despite the similar student outcomes produced by the two course formats, students in the hybrid sections reported considerably lower satisfaction with their experience. Thus, these findings suggest that online learning technologies alone, at least in their current form, are not a panacea for higher education's challenges. Students place high value on their personal experiences with faculty. If students took many courses in a hybrid format, the lower satisfaction could lead to reduced retention and persistence rates. Efforts to expand the use of technology-enhanced education on traditional college campuses will thus have to address both the micro issues around designing high-quality content and the macro issues of integrating these technology-enhanced courses into a rich curriculum that still makes use of the interpersonal interactions at the core of a brick-and-mortar campus environment.

## Notes

1. Example providers include OLI and commercial publishers such as Pearson and McGraw Hill.
2. This article primarily reports the results of quantitative comparisons of hybrid and traditional instruction. The results of the full study, including quantitative and qualitative components, are reported in Griffiths et al. (2014).
3. Udacity and San Jose State partnered to use a MOOC in remedial math courses. They assessed the results (which they deemed unsatisfactory) by comparing their pass rates with previous years. The lack of control variables and common assessments makes it difficult to assess the usefulness of this comparison (Firmin, Schiorring, Whitmer, Willett, & Sujitparapitaya, 2013).
4. We also set up 10 case studies of smaller courses using MOOCs to understand the implementation challenges of working with a MOOC in a wider variety of courses. These case studies are discussed in Griffiths et al. (2014).
5. In some cases, individual instructors submitted statements, and in others, a coordinator at the institutions collected proposals and put forward the strongest candidates for consideration.
6. In general, the instructors volunteered to participate, but in some cases, they were volunteered by a course coordinator. Institutions were offered $20,000 for side-by-side tests to cover course buyout or summer planning time for faculty members who participated in the test, time for instructors to attend USM workshops, administrative costs for preparing data, and any other costs incurred in organizing the tests. The $20,000 covered all people involved in the study, and in all cases, it was divided among multiple parties. Each institution presented a budget to USM explaining how the funds would be spent. These incentives may have affected who participated in the study. It is unlikely that faculty would have participated without these incentives because of how much time it took to create these courses. The stipends were slightly higher than others offered by USM for course redesign because of the research-related activities.
7. Two courses, sociology and statistics for sociology, were initially included in the planned set of side-by-side tests, but they were excluded from the analyses in this article because one faculty member was unable to secure permission to use the MOOC and another professor decided to create his own videos after using a few from the MOOC.
8. Instructors with courses that were not sufficiently large or that could not create control sections participated in the project as case studies. These case studies included a number of smaller and upper-level courses (Griffiths et al., 2014).
9. We were unable to impose strict requirements about class meeting time and the number and length of videos to be used because each instructor felt that even minor tweaks to our guidelines could improve the course. In addition, the variance in content and format of the MOOCs and OLI made it difficult to require each course to use them in the exact same way.

10. One course taught all of the materials in class, and the MOOC videos were only a supplement.
11. Six of the seven courses had control sections that were traditional lectures. One course's control section was a hybrid, but it did not include OLI.
12. The research team relied on instructors to determine what students should learn in their courses and to develop appropriate assessments.
13. Parental income was missing for students who did not submit a Free Application for Federal Student Aid or answer the question about parental income on the student survey. ACT scores were converted to the SAT scale for students missing SAT scores. Computer science and biology B did not offer pretests. Excluding these courses, only 11% of students were missing pretest scores. These scores were missing for students who were absent on the day the pretest was administered and did not retake it.
14. One group of instructors could not agree on a common assessment, so this course was excluded from the analysis of posttest/final exam results.
15. Copies of both survey instruments are available in Griffiths et al. (2014).
16. For the initial survey, the response rates were 75% and 79% for the treatment and control groups, respectively, and for the final survey, they were 76% and 75%, respectively.
17. We do not know how advisors may have influenced student placements. It is possible that advisors dissuaded certain students from taking certain course formats. Advisors would have known which courses were hybrid and traditional formats based on the course listing. However, they would not have known the details of the hybrid courses and which materials were being used unless they asked the instructors.
18. Pretest scores were somewhat different, in favor of the treatment group, but we did not attach too much importance to this finding given the difference in test-taking rates (with less missing data in the treatment group). We return to this issue in the Robustness Checks section.
19. Institutions were unable to provide data on students switching courses for students who changed their enrollment prior to the drop deadline. This deadline varied by campus but was usually in the middle of the semester. Any student who withdrew from the course after this deadline was captured in our data.
20. The instructors of the control sections had more experience teaching with technology, but this experience does not necessarily refer to the courses being studied. If these instructors incorporated technology in their existing courses, it was not in the same way or to the same extent as in the treatment sections. In most cases, control instructors' reference to technology indicated that they used a course management site or a platform with homework problems.
21. We obtained similar results using a probit model. One reason we used linear probability models instead of probit for the nonbinary outcomes is that probit drops observations where one of the covariates perfectly predicts the outcome.
22. One limitation of the posttest and finals is that some of them were multiple-choice. Thus, they may not have captured some aspects of student learning as well as tests that included a mix of multiple-choice and open-response questions.
23. Three courses had common finals, two had a subset of common exam questions, and one course had a common midterm exam (because the MOOC was only used in the first half of the semester). The instructors for biology B were unable to agree on a common final assessment.
24. The results in Table 4 adjust for course-specific differences in performance. In other words, they show a weighted average of within-course differences between the hybrid and traditional sections across all seven tests. The raw data are roughly similar, with

pass rates of 83% and 86% in the traditional and hybrid sections, respectively, and final assessment scores of 70% and 70%, respectively.

25. These results are available from the authors upon request. The only notable exception to the similarity in results was the estimated treatment–control difference for pass rates in the statistics course, which was .127 (and statistically significant) without controls and .045 (but not statistically significant) with controls.
26. These results are available upon request.
27. These were the MOOCs that had been run at least once before, had been well received by public audiences, and did not run into major pedagogical or technical problems.



## References

Allen, I. E., & Seaman, J. (2014). *Grade change: Tracking online education in the United States*. Babson Park, MA: Babson Survey Research Group and Quahog Research Group. Retrieved from http://www.onlinelearningsurvey.com/reports/gradelevel.pdf

Bailey, T. R., Jaggars, S. S., & Jenkins, D. (2015). *Redesigning America's community colleges: A clearer path to student success*. Cambridge, MA: Harvard University Press.

Bishop, J. L., & Verleger, M. A. (2013). The flipped classroom: A survey of the research. *ASEE National Conference Proceedings, June 2016*.

Borden, J. (2014). MOOCs are dead—Long live the MOOC. *WIRED*. Retrieved from https://www.wired.com/insights/2014/08/moocs-are-dead-long-live-the-mooc/

Bound, L., Lovenheim, M., & Turner, S. (2012). Increasing time to baccalaureate degree in the United States. *Education Finance and Policy*, *7*, 375–424. doi:10.1162/EDFP_a_00074

Bowen, W. G. (2013). *Higher education in the digital age*. Princeton, NJ: Princeton University Press.

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*, 94–111. doi:10.1002/pam.21728

Brown, D. G. (2001). Hybrid courses are best. *Syllabus Magazine*, *15*(3), 22.

Burns, K., Duncan, M., Sweeney, D. C. I. I., North, J. W., & Ellegood, W. A. (2013). A longitudinal comparison of course delivery modes of an introductory information systems courses and their impact on a subsequent information systems course. *MERLOT Journal of Online Learning and Teaching*, *9*, 453–467.

Chenoweth, N. A., Ushida, E., & Murday, K. (2006). Student learning in hybrid French and Spanish courses: An overview of language online. *Calico Journal*, *24*, 115–146.

Christensen, G., Steinmetz, A., Alcorn, B., Bennett, A., Woods, D., & Emanuel, E. J. (2013). *The MOOC phenomenon: Who takes massive open online courses and why?* Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2350964

Coursera. (2016, November 10). *Meet our partners*. Retrieved from https://www.coursera.org/about/partners

Educause Library. (2015) *Massive open online courses*. Retrieved from http://www.educause.edu/library/massive-open-online-course-mooc

edX. (2015) *About edX*. Retrieved from http://thux.thu.edu.tw/about

Fain, P. (2012, November 14). Establishment opens door for MOOCs. *Inside Higher Ed*. http://www.insidehighered.com/news/2012/11/14/gates-foundation-and-ace-go-big-mooc-related-grants#sthash.hZ5LIsgV.Epr66KXO.dpbs

Figlio, D., Rush, M., & Yin, L. (2013). Is it live or is it Internet? Experimental estimates of the effects of online instruction on student learning. *Journal of Labor Economics*, *31*, 763–784. doi:10.1086/669930

Firmin, R., Schiorring, E., Whitmer, J., Willett, T., & Sujitparapitaya, S. (2013). *Preliminary summary SJSU augmented online learning environment pilot project*. San Jose, CA: The Research & Planning Group for California Community Colleges. Retrieved from http://www.sjsu.edu/chemistry/People/Faculty/Collins_Research_Page/AOLE%20Report%20-September%2010%202013%20final.pdf

Griffiths, R., Chingos, M., Mulhern, C., & Spies, R. (2014, July 10). Interactive online learning on campus: Testing MOOCs and other platforms in hybrid formats in the University System of Maryland. *Ithaka S+R*. Retrieved from http://www.sr.ithaka.org/research-publications/interactive-online-learning-on-campus

Joyce, T. J., Crockett, S., Jaeger, D. A., Altindag, O., & O'Connell, S. D. (2014). *Does classroom time matter? A randomized field experiment of hybrid and traditional lecture formats in economics* (No. w20006). Cambridge, MA: National Bureau of Economic Research.

Kwak, D. W., Menezes, F. M., & Sherwood, C. (2014). Assessing the impact of blended learning on student performance. *Economic Record*, *91*, 91–106. doi:10.1111/1475-4932.12155

Lack, K. A. (2013, March 21). Current status of research on online learning in postsecondary education. *Ithaka S+R*. Retrieved from http://www.sr.ithaka.org/publications/current-status-of-research-on-online-learning-in-postsecondary-education

Mazoue, J. G. (2013, January 28). The MOOC model: Challenging traditional education. *Educause Review Online*. Retrieved from http://www.educause.edu/ero/article/mooc-model-challenging-traditional-education

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development.

Mulhern, C., Spies, R. R., Staiger, M. P., & Wu, D. D. (2015, March 4). The effects of rising student costs in higher education: Evidence from public institutions in Virginia. *Ithaka S+R*. Retrieved from http://www.sr.ithaka.org/publications/the-effects-of-rising-student-costs-in-higher-education

Oblender, T. E. (2002). A hybrid course model: One solution to the high online drop-out rate. *Learning & Leading with Technology*, *29*(6), 42–46.

Open Learning Initiative. (2015). *Our key goals*. Retrieved from http://oli.cmu.edu/get-to-know-oli/learn-more-about-oli

Pappano, L. (2012, November 2). The year of the MOOC. *The New York Times*. Retrieved from http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html?_r=0

Rivard, R. (2013, March 22). Coursera's contractual elitism. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2013/03/22/coursera-commits-admitting-only-elite-universities

Shah, D. (2015, December 21). By the numbers: MOOCs in 2015. *Class Central*. Retrieved from https://www.class-central.com/report/moocs-2015-stats/

Stokes, P. (2013, February 22). The particle accelerator of learning. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/views/2013/02/22/look-inside-edxs-learning-laboratory-essay

Tanyel, F., & Griffin, J. (2014). *A ten-year comparison of outcomes and persistence rates in online versus face-to-face courses*. Retrieved from http://www.westga.edu/~bquest/2014/onlinecourses2014.pdf

Verhoeven, P., & Rudchenko, T. (2013). Student performance in a principle of microeconomics course under hybrid and face-to-face delivery. *American Journal of Educational Research*, *1*, 413–418. doi:10.12691/education-1-10-1

The White House. (2015) *Higher education*. Retrieved from https://www.whitehouse.gov/issues/education/higher-education

Xu, D., & Jaggars, S. S. (2011). The effectiveness of distance education across Virginia's community colleges: Evidence from introductory college-level math and English courses. *Educational Evaluation and Policy Analysis*, *33*, 360–377. doi:10.3102/0162373711413814

Xu, D., & Jaggars, S. S. (2013). The impact of online learning on students' course outcomes: Evidence from a large community and technical college system. *Economics of Education Review*, *37*, 46–57. doi:10.1016/j.econedurev.2013.08.001

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*, 633–659. doi:10.1353/jhe.2014.0028



**Table A1.** Robustness Checks and Additional Analyses.

| | Pass | Posttest/Final | Grade |
|---|---|---|---|
| Main results | .030 | .024* | −.024 |
| | (.032) | (.012) | (.169) |
| | *1,594* | *1,187* | *1,564* |
| Add instructor controls | .037 | .039*** | −.017 |
| | (.028) | (.012) | (.165) |
| | *1,554* | *1,152* | *1,524* |
| Pretest control | .009 | .017 | −.107 |
| | (.036) | (.013) | (.178) |
| | *1,594* | *1,187* | *1,564* |

*Note*. * $p < .1$. *** $p < .01$. Robust standard errors adjusted for clustering by section appear in parentheses. Number of observations appears in italics. All models include dummies for each course. All models include student controls, including SAT math and verbal scores, cumulative grade point average at the beginning of the semester, race/ethnicity, gender, parents' income, and age. Missing values are identified by dummies for each variable, with continuous variables imputed using an arbitrary value (in this case, 0). Instructor controls include tenure-track status, section size, and years of teaching experience.

**Table A2.** Differences between outcomes in hybrid and traditional sections, by course.

| | Bio A | Bio B | Bio C | Comms | CompSci | Precalculus | Stats |
|---|---|---|---|---|---|---|---|
| Hybrid effect on pass rates | 0.200*** | −0.039 | −0.043 | 0.038 | −0.068 | 0.012 | 0.045 |
| | (0.052) | (0.033) | (0.027) | (0.041) | (0.070) | (0.090) | (0.054) |
| Observations | 288 | 168 | 478 | 207 | 172 | 122 | 159 |
| Control mean | 0.65 | 0.98 | 0.91 | 0.89 | 0.73 | 0.78 | 0.78 |
| Hybrid effect on posttest/final | 0.052*** | | 0.023 | −0.003 | −0.033 | 0.101*** | −0.021 |
| | (0.013) | | (0.015) | (0.019) | (0.031) | (0.030) | (0.021) |
| Observations | 231 | | 455 | 139 | 119 | 97 | 146 |
| Control mean | 0.63 | | 0.72 | 0.89 | 0.71 | 0.67 | 0.68 |
| Hybrid effect on course grade | 0.755*** | −0.667*** | −0.181** | 0.164 | −0.249 | −0.408 | −0.206 |
| | (0.131) | (0.154) | (0.085) | (0.169) | (0.232) | (0.278) | (0.170) |
| Observations | 287 | 142 | 478 | 207 | 172 | 119 | 159 |
| Control mean | 1.29 | 3.01 | 2.27 | 3.11 | 1.98 | 2.24 | 2.06 |
| Technology | OLI | OLI | OLI | MOOC | MOOC | MOOC | MOOC |
| Average size of treatment section | 24 | 29 | 120 | 26 | 23 | 28 | 32 |
| Average size of control section | 147 | 40 | 120 | 15 | 17 | 22 | 32 |
| Minutes per week treatment section met | 50 | 50 | 50 | 80 | 75–100 | 120 | 150 |
| Minutes per week control section met | 150 | 150 | 50 | 160 | 150 | 220 | 150 |
| Aware of format enrolled in? | Yes | Yes | No | No | No | Yes | No |
| Additional notes | The control instructor was known to have difficult grading standards. | The instructors were unable to agree on a common assessment. | The control section was also a hybrid. | | | | The MOOC materials were supplemental.. |

*Note*. OLI = Open Learning Initiative; MOOC = massively open online course. ** $p < .05$. *** $p < .01$. Robust standard errors appear in parentheses. All models include controls for SAT math and verbal scores, cumulative grade point average at the beginning of the semester, race/ethnicity, gender, parents' income, and age. Missing values are identified by dummies for each variable, with continuous variables imputed using an arbitrary value (in this case, 0).