ITHAKA S+R

Interactive Learning Online at Public Universities: Evidence from Randomized Trials

William G. Bowen
Matthew M. Chingos
Kelly A. Lack
Thomas I. Nygren

May 22, 2012

Ithaka S+R is a strategic consulting and research service provided by ITHAKA, a not-for-profit organization dedicated to helping the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways. Ithaka S+R focuses on the transformation of scholarship and teaching in an online environment, with the goal of identifying the critical issues facing our community and acting as a catalyst for change. JSTOR, a research and learning platform, and Portico, a digital preservation service, are also part of ITHAKA.

Preface

Higher education is facing serious challenges in the United States. There is increasing concern about rising costs, the quality of education, and that the nation is losing its "competitive edge." Online learning—specifically highly interactive, closed-loop, online learning systems that we call ILO or Interactive Learning Online—holds the promise of broadening access to higher education to more individuals, while also lowering costs for students. But is the quality there?

In our first report in this area, "Barriers to Adoption of Online Learning Systems in U. S. Higher Education," we highlighted a broad, widely held concern about the quality of learning outcomes achieved through online learning. But do we actually know how interactive online learning systems really compare to the in-classroom experience? This second report was designed to help find answers.

We used a strictly quantitative methodology to compare the two learning approaches in a rigorous way. In six different public institutions, we arranged for the same introductory statistics course to be taught. In each instance, a "control" group was enrolled in a traditional classroom-based course; then, a "treatment" group took a hybrid course using a prototype machine-guided mode of instruction developed at Carnegie Mellon University in concert with one face-to-face meeting each week. Students were assigned to these two groups by means of a carefully designed randomization methodology. The research we conducted was designed to answer these questions:

- Can sophisticated, interactive online courses be used to maintain or improve basic learning outcomes (mastery of course content, completion rates, and time-to-degree) in introductory courses in basic subjects such as statistics?
- Are these courses as effective, or possibly more effective, for minority and low-socioeconomic-status students and for other groups subject to stereotype threat? Or, are these groups less well suited to an online approach?
- Are such courses equally effective with not-so-well-prepared students and well-prepared students?

The results of this study are remarkable; they show comparable learning outcomes for this basic course, with a promise of cost savings and productivity gains over time.

More research is needed. Even though the analysis was rigorous, it was a single course. We need to learn more about the adaptability of existing platforms for offering other courses in different environments. Ithaka S+R is committed to continuing this research and sharing our findings broadly.

We look forward to continuing to engage with all those who care about higher education to help deliver on the potential that new technologies provide.

DEANNA MARCUM
Deanna.Marcum@ithaka.org
Managing Director, Ithaka S+R

Interactive Learning Online at Public Universities: Evidence from Randomized Trials

Interactive Learning Online at Public Universities: Evidence from Randomized Trials

William G. Bowen, Matthew M. Chingos, Kelly A. Lack, and Thomas I. Nygren[1]
May 22, 2012

Introduction

The topic of online learning in higher education is of obvious importance. The serious economic and social problems facing the U.S.—high unemployment, slow growth, and severe inequalities—are related, many believe, to failures of the

1 The authors are all associated with Ithaka S+R (the Strategy and Research arm of ITHAKA), which sponsored this study. Bowen is a senior advisor to Ithaka S+R, Chingos is a senior research consultant at Ithaka S+R and a fellow at the Brookings Institution's Brown Center on Education Policy, Lack is a research analyst, and Nygren is a project director and senior business analyst for Ithaka S+R. The authors wish to thank the foundations that supported this work: the Carnegie Corporation of New York, the William and Flora Hewlett Foundation, the Spencer Foundation, and a fourth foundation that has asked to remain anonymous. We also thank our colleagues at ITHAKA—and Larry Bacow, Johanna Brownell, Jackie Ewenstein, and Kevin Guthrie in particular—for their generous help all along the way. But most of all, we wish to thank our faithful friends on the participating campuses for their hard work and patience with us; their names are appended to this report. A number of these individuals (as well as others) have commented on a draft of the report, but the authors are, of course, fully responsible for the views expressed here and for any errors that remain.

Ithaka S+R has sponsored three studies of online learning, of which this is the longest lasting. The two other studies are now available on the Ithaka S+R website. See "Barriers to Adoption of Online Learning Systems in U.S. Higher Education" by Lawrence S. Bacow, William G. Bowen, Kevin M. Guthrie, Kelly A. Lack, and Matthew P. Long, and "Current Status of Research on Online Learning in Postsecondary Education" by William G. Bowen and Kelly A. Lack (both available online at http://www.sr.ithaka.org/).

Levels of educational attainment in this country have been stagnant for almost three decades, while many other countries have been making great progress in educating larger numbers of their citizens.

U.S. education system, including higher education.[2] Levels of educational attainment in this country have been stagnant for almost three decades, while many other countries have been making great progress in educating larger numbers of their citizens. There is growing concern that the U.S. is losing its "competitive edge" in an increasingly knowledge-driven world. Also, substantial achievement gaps related to race and socioeconomic status persist and have a great deal to do with worrying "inequities." Moreover, there are good reasons to believe that these two problems are closely related.[3]

The Cost Squeeze in Higher Education

At the same time, higher education, especially in the public sector, is increasingly short of resources. States continue to cut back appropriations in the face of fiscal constraints and pressures to spend more on other things, such as health care and retirement expenses.[4] California is a dramatic case in point. Lack of funding has caused California colleges and universities to reduce the size of their entering classes at the very time when increasing numbers of students are seeking to enroll.[5] Higher tuition revenues might be an escape valve, but there is great concern about tuition levels and increasing resentment among students and their families that is having political reverberations. President Obama, in his

2 The authors agree that there is an important connection between educational outcomes and the economic performance of a country. But we would warn against exaggerating the power of the connection. In the case of the U.S., for example, the recent recession and the slow rate of growth seen in the last few years surely owe more to the 2008 financial excesses than they do to deficiencies in the country's higher education system. As Jacob Weisberg pointed out in Newsweek in 2010 with respect to the recent recession, "there are no strong candidates for… a single factor that would have caused the crisis in the absence of any others" (Weisberg's piece can be found online at http://www.thedailybeast.com/newsweek/2010/01/08/what-caused-the-great-recession.html).

3 See *Equity and Excellence in American Higher Education* by William G. Bowen, Martin A. Kurzweil, and Eugene M. Tobin (2005) for an extended discussion of the historical record and of the likely connections, going forward, between achievement gaps and overall levels of educational attainment. See also David Leonhardt's October 8, 2011 column in the New York Times, "The Depression: If Only Things Were That Good," in which he argues that the U.S. is worse off today than it was in the 1930s because innovation is lagging—which he attributes in no small part to deficiencies in education (http://www.nytimes.com/2011/10/09/sunday-review/the-depression-if-only-things-were-that-good.html?_r=1&pagewanted=all.) Of course, lagging rates of educational attainment have their origins in low high school graduation rates. See Henry M. Levin and Cecilia E. Rouse, "The True Cost of High School Dropout," New York Times, January 25, 2012. (http://www.nytimes.com/2012/01/26/opinion/the-true-cost-of-high-school-dropouts.html). But these problems are then compounded by low completion rates among those who both graduate from high school and enter college; see *Crossing the Finish Line: Completing College at America's Public Universities* (2009) by William G. Bowen, Matthew M. Chingos, and Michael S. McPherson.

4 A report released in spring 2012 by the State Higher Education Executive Officers, entitled "State Higher Education Finance FY 2011" (http://www.sheeo.org/finance/shef/SHEF_FY2011-EARLY_RELEASE.pdf), documents the dire economic circumstances of many public institutions.

5 In November 2008, California State University became the first public university to limit enrollment when, despite a 20% increase in applications from prospective first-year students, it decided to reduce its student body by 10,000 students, following a $200 million decrease in tax revenue that academic year coupled with an additional $66 million cut (see "Under Financial Stress, More Colleges Cap Enrollments" (November 26, 2008) in TIME, http://www.time.com/time/nation/article/0,8599,1861861,00.html). The University of California and California Community College systems have since followed suit in the face of limited funding available from the state (see the August 5, 2009 article "Budget cuts devastate California higher education" in The Washington Examiner, http://washingtonexaminer.com/science-and-technology/2009/08/budget-cuts-devastate-california-higher-education).

Higher education, especially in the public sector, is increasingly short of resources. States continue to cut back appropriations in the face of fiscal constraints and pressures to spend more on other things, such as health care and retirement expenses.

2012 State of the Union address and in subsequent speeches, has decried rising tuitions, called upon colleges and universities to control costs, and proposed to withhold access to some Federal programs for colleges and universities that did not address "affordability" issues or meet completion tests.[6]

Today, a variety of higher education institutions must confront the challenge of how to manage costs in the face of tighter funding. While the proportion of education spending drawn from tuition revenues rose across all institutions, increases in tuition often outpaced increases in education and related spending (i.e. spending on instruction, student services, and some support and maintenance costs related to these functions), calling into question the sustainability of the current funding model.[7] Moreover, the first survey of provosts and chief academic officers by *Inside Higher Ed* found that on the question of institutional effectiveness in controlling costs, "over 15 percent of all provosts gave their institutions marks of 1 or 2 on effectiveness [on a scale from 1 to 7, with 7 being very effective]."[8] It is equally noteworthy that *very* few chief academic officers (and especially those at both public and private doctoral universities) gave their institutions high marks on this metric. Recognition of the problem is widespread; "solutions" have been hard to come by.

A fundamental source of the problem is the "cost disease," based on the handicraft nature of education with its attendant lack of opportunities for gains in productivity, which one of the authors of this report (Bowen) promulgated in the 1960s, in collaboration with William J. Baumol. But the time may (finally!) be at hand when advances in information technology will permit, under the right circumstances, increases in productivity that can be translated into reductions in

6 See "Remarks by the President in State of the Union Address," January 24, 2012 (transcript available at (http://www.whitehouse.gov/the-press-office/2012/01/24/remarks-president-state-union-address). Three days later, Obama spoke about college affordability at the University of Michigan (transcript available at http://www.whitehouse.gov/the-press-office/2012/01/27/remarks-president-college-affordability-ann-arbor-michigan). This speech does not, however, contain more details concerning how "affordability" is to be measured or what penalties are to be imposed on those who fail to pass the requisite tests. As Molly Broad, president of the American Council on Education, said after the speech: "The devil is in the [unspecified] details" ("Mixed Reviews of Obama Plan to Keep Down College Costs," January 28, 2012, New York Times, http://www.nytimes.com/2012/01/28/education/obamas-plan-to-control-college-costs-gets-mixed-reviews.html).

7 According to the College Board's 2011 Trends in College Pricing Report (http://trends.collegeboard.org/downloads/College_Pricing_2011.pdf), tuition at public two-year universities increased, on average, by 8.7% relative to the 2010-2011 academic year, and tuition at public four-year institutions for the 2011-2012 academic year increased, on average, by 8.3% for instate students and by 5.7% for out of state students. In keeping with the trend over the previous four years, students attending private institutions experienced smaller percentage increases (4.5% for private not-for-profit four-year institutions and 3.2% for private for-profit institutions).

8 See Scott Jaschik, "Mixed Grades: A Survey of Provosts," Inside Higher Education, January 25, 2012, http://www.insidehighered.com/news/survey/mixed-grades-survey-provosts.

There are also concerns that at least some kinds of online learning are low quality and that online learning in general de-personalizes education. In this regard, it is critically important to recognize issues of nomenclature: "online learning" is hardly one thing. It comes in a dizzying variety of flavors.

the cost of instruction.[9] Greater—and smarter—use of technology in teaching is widely seen as a promising way of controlling costs while also reducing achievement gaps and improving access. The exploding growth in online learning is often cited as evidence that, at last, technology may offer pathways to progress.[10] Online learning is seen by a growing number of people as a way of breaking free of century-old rigidities in educational systems that we have inherited. The much-discussed book on disruptive technologies and universities by Clayton Christensen and Henry Eyring is perhaps the best example of the attention being given to online technologies as a way of changing profoundly the way we educate students.[11]

There are, however, also concerns that at least some kinds of online learning are low quality and that online learning in general de-personalizes education. *In this regard, it is critically important to recognize issues of nomenclature: "online learning" is hardly one thing.* It comes in a dizzying variety of flavors, ranging from simply videotaping lectures and posting them for any-time access, to uploading materials such as syllabi, homework assignments, and tests to the Internet, all the way to highly sophisticated interactive learning systems that use cognitive tutors and take advantage of multiple feedback loops. The varieties of online learning can be used to teach many kinds of subjects to different populations in diverse institutional

9 Bowen's co-author in the promulgation of the "cost disease," William J. Baumol, has continued to discuss its relevance not only for education but also for sectors such as the performing arts and heath care. For the initial statement of this proposition, see William J. Baumol and William G. Bowen, *Performing Arts: The Economic Dilemma, Twentieth Century Fund* (1968). In essence, the argument is that in fields such as the performing arts and education, there is less opportunity than in other fields to improve productivity (by, for example, substituting capital for labor), that unit labor costs will therefore rise inexorably as these sectors have to compete for labor with other sectors in which productivity gains are easier to come by, and that the relative costs of labor-intensive activities such as chamber music and teaching will therefore continue to rise. As Bowen argued in his Romanes lecture, for a number of years advances in information technology have in fact increased productivity, but these increases have been enjoyed primarily in the form of more output (especially in research) and have generally led to higher, not lower, total costs. (For the text of the Romanes lecture, see William G. Bowen, "At a Slight Angle to the Universe: The University in a Digitized, Commercialized Age," Princeton University Press, 2001; the text is also available on the Andrew W. Mellon Foundation website: http://www.mellon.org/internet/news_publications/publications/romanes.pdf.)

10 A November 2011 report by the Sloan Consortium and the Babson Survey Research Group shows that between fall 2002 and fall 2010, enrollments in online courses increased much more quickly than total enrollments in higher education. During this time period, the number of online course enrollments grew from 1.6 million to 6.1 million, amounting to a compound annual rate of 18.3% (compared with a rate of 2% for course enrollments in general)—although between fall 2009 and fall 2010 online enrollments grew more slowly, at 10.1%. More than three of every 10 students in higher education now take at least one course online. In addition to the growth in what we call "online" or "hybrid" courses—however nebulous that terminology may be—we also "feel" the pervasiveness of the Internet in higher education by the increasing use of it in the form of course management systems or virtual reading materials/electronic textbooks incorporated into the curriculum. Even courses that are called "traditional" almost always involve some use of digital resources.

11 See Clayton M. Christensen, and Henry J. Eyring, *The Innovative University: Changing the DNA of Higher Education from the Inside Out*, San Francisco: Jossey-Bass, 2011. An October 2, 2011 New York Times op-ed piece by Bill Keller, aptly titled "The University of Wherever," is another illustration of the high visibility and high stakes of the debate over online education (http://www.nytimes.com/2011/10/03/opinion/the-university-of-wherever.html?pagewanted=all).

settings. A key point, if an obvious one, is that there is no one approach that is right for every student or every setting. In important respects, the online learning marketplace reflects the diversity of American higher education itself.[12]

As resistant as some may still be even to think about seeking productivity gains in order to reduce teaching costs, there is simply no denying the need to look more closely than ever before at the relation between certain "outputs" (approximated, for example, by degrees conferred) and "inputs" (the mix of labor and capital that defines educational production functions).[13] It is essential that the limited resources available to higher education be used as effectively as possible. For these reasons, the research reported here is concerned with both educational outcomes and costs, seen as two blades of the scissors.

Organization of This Report

The next section of this report describes a two-year effort we have made to test rigorously the learning outcomes achieved by a prototype interactive learning online course delivered in a hybrid mode (with some face-to-face instruction) on public university campuses in the Northeast and Mid-Atlantic. Before presenting our findings, we devote space to explaining our randomization methodology—both because the findings can only be understood against the backdrop of the methodology and because the research design may be of independent interest to some readers.[14] This section—which contains the results of the main part of our research—is followed by a briefer discussion of the potential cost savings that can conceivably be achieved by the adoption of hybrid-format online learning systems. We explain why we favor using a cost simulation approach to estimate potential savings, but we relegate to Appendix B the highly provisional results we obtained by employing one set of assumptions in a cost simulation model. We end the main body of the report with a short conclusion that considers barriers to the adoption of online learning systems that are truly interactive, steps that might be taken to overcome such barriers, and the need to take a system-wide perspective in addressing these extremely important issues.

12 As Henry Bienen (president emeritus of Northwestern and chairman of the board of Rasmussen College, a for-profit university, as well as chairman of ITHAKA) points out, for many institutions seeking to address the needs of adult learners and others who are not candidates for places in traditional colleges and universities, there is no choice: online education, in some form, is the only way that many people can acquire more skills and earn a college degree, the returns on which have skyrocketed in the past three decades. But online education is also increasingly common in colleges and universities that educate "traditional" students. It is seen as a "revenue-generating" force in many institutions, both four-year and two-year and both public and private. See "Barriers to Adoption of Online Learning Systems in U.S. Higher Education" by Bacow et al.

13 Some argue—and we heartily agree—that the "output" of higher education has broader dimensions and includes both research results and also the contribution that the entire system of higher education makes to the effective functioning of a democratic society. But it will not do to allow emphasis on these larger (and hard-to-measure) contributions to obscure the need to look carefully, and with a somewhat skeptical eye, at how effectively institutions utilize resources to achieve straightforward aims such as improving graduation rates.

14 Readers interested in methodology may be especially interested in Appendix C to this report, which contains a detailed discussion of "lessons learned" from our experience in carrying out this complicated research project. We wish only that we had had access to this recitation of what to do and what not to do before we started on this adventure! We learned many of these lessons "the hard way."

The first and most ambitious part of our research was directed at assessing the educational outcomes associated with what we term "interactive learning online" or "ILO." By "ILO" we refer to highly sophisticated, interactive online courses in which machine-guided instruction can substitute for some (though not usually all) traditional, face-to-face instruction. Course systems of this type take advantage of data collected from large numbers of students in order to offer each student customized instruction, as well as allow instructors to track students' progress in detail so that they can provide their students with more targeted and effective guidance. As several leaders of higher education made clear to us in preliminary conversations, absent real evidence about learning outcomes there is no possibility of persuading most traditional colleges and universities, and especially those regarded as thought leaders, to push hard for the introduction of ILO technologies that begin to substitute machine-guided instruction for traditional forms of teaching in appropriate settings.

The most ambitious part of our research was directed at assessing the educational outcomes associated with what we term "interactive learning online" or "ILO." By "ILO" we refer to highly sophisticated, interactive online courses in which machine-guided instruction can substitute for some (though not usually all) traditional, face-to-face instruction.

We set out to provide at least tentative answers to these questions:

- Can sophisticated, interactive online courses be used to maintain or improve basic learning outcomes (mastery of course content, completion rates, and time-to-degree)?
- Are these courses as effective, or possibly more effective, for minority and low-socioeconomic-status students and for other groups subject to stereotype threat?
- Are they equally effective with not-so-well-prepared students and well-prepared students?
- Are they equally effective in a variety of campus settings—community colleges versus four-year colleges, commuter colleges versus colleges with more students in residence?

Research Design

In thinking about research design, we began by looking closely at existing research. There have been literally thousands of studies of "online learning," but unfortunately the great majority are deficient in one way or another—often for reasons beyond the control of the principal investigators.[15] Very few look directly at the teaching of large introductory courses in basic fields at major public universities, where the great majority of undergraduate students pursue either associate or baccalaureate degrees, presumably because very few ILO courses have been

15 A detailed summary of existing research has been compiled by our staff (especially Lack); but it is too lengthy to include here. See "Current Status of Research on Online Learning in Postsecondary Education" by Bowen and Lack.

offered in these settings.[16] Very few of the studies use randomized assignment techniques to create "treatment" and "control" groups that can be used to reduce otherwise ubiquitous selection effects that make it hard to interpret findings.

To overcome these limitations, we decided to work with seven instances of a prototype ILO statistics course at six public university campuses (including two separate courses in two departments on one campus) that agreed to cooperate in a carefully designed research project utilizing random assignment techniques. Two of these campuses are part of the State University of New York (SUNY); two are part of the University of Maryland; and two are part of the City University of New York (CUNY). The individual campuses involved in this study were, from SUNY, the University at Albany and SUNY Institute of Technology; from the University of Maryland, the University of Maryland, Baltimore County and Towson University; and, from CUNY, Baruch College and City College. The seven courses, with their fall 2011 enrollments, are shown in Table 1.

We also attempted to include three community colleges in New York and Maryland. We were ultimately unable to include data from these campuses in our study for several reasons. At one of the three community colleges, multiple changes in leadership compromised the implementation of the randomized research protocol. At the second community college, a large number of study participants never took the course, and among those who did, almost a quarter switched into a format different from the one to which they were randomly assigned. Additionally, data on final exam and standardized test scores were unavailable for a substantial proportion of this campus' study participants. At the third community college, much of the data were provided too late to incorporate into our primary analysis. We strongly caution readers against assuming that the findings reported here for four-year colleges necessarily apply to community colleges. Vigorous efforts notwithstanding, we were unable to obtain hard evidence on this key question.

16 Our focus on students attending public institutions is not meant to denigrate the importance of either the private non-profit sector or the for-profit sector. Nor is it meant to denigrate professional programs aimed at working adults. But it is the public colleges and universities, which educate more than three-quarters of undergraduates at degree-granting institutions (according to the College Board's 2011 report, cited above), that face the most consequential challenges in raising attainment rates and closing achievement gaps while simultaneously reducing costs and restraining tuition increases.

TABLE 1. PARTICIPATING COURSES/INSTITUTIONS, FALL 2011

	Course Enrollment	Study Participants
Institution A	850	97
Institution B	877	229
Institution C	235	92
Institution D	86	16
Institution E, Department 1	337	31
Institution E, Department 2	473	50
Institution F	188	90
Total	3,046	605

Notes: Study participants are students who consented to be in our study and were randomly assigned to a traditional or hybrid format of the introductory statistics class.

The population of institutions and students in the study is both large enough and diverse enough to allow us to explore most of the questions listed above in the context of four-year public institutions.

We do not claim that these six campuses are a statistically valid sample of even public higher education, never mind all of higher education. But this set of six does include: (a) major urban universities with large commuting populations of students, as well as universities with more residential students; and (b) large numbers of minority students and students from low-socioeconomic-status families (as shown in Tables 2 and 3). Thus, the population of institutions and students in the study is both large enough and diverse enough to allow us to explore most of the questions listed above in the context of four-year public institutions.

More specifically, this research was designed to test as rigorously as possible the learning effectiveness of a particular interactive statistics course developed at Carnegie Mellon University (CMU)—viewed as a prototype of other ILO coures.[17] While the CMU course can be delivered in a fully online environment, in this study it was used in a "hybrid" mode in which most of the instruction was delivered through the interactive online materials, but the online instruction was supplemented by a one-hour-per-week face-to-face session in which students could ask questions or be given targeted assistance.

The exact research protocol varied by campus in accordance with local policies, practices, and preferences, and we describe these protocols in detail in Appendix Table A1, and on Ithaka S+R's website where there is a narrative description; Appendix Table A1 also presents summary data on enrollments and section sizes in each format (often the hybrid-format sections were somewhat smaller than the traditional-format sections). The general procedure followed was: 1) at or before the beginning of the semester, students registered for the introductory statistics course were asked to participate in our study, and modest incentives were offered;[18] 2) students who consented to participate filled out a baseline survey; 3) study participants were randomly assigned to take the class in a traditional or hybrid format; 4) study participants were asked to take the CAOS test of statistical literacy[19] at the beginning of the semester; and 5) at the end of the semester,

17 We prefer the "ILO" acronym to others, including the "OLI" acronym used by CMU to stand for "Open Learning Initiative." The term "ILO"—for interactive learning online—is not specific to CMU's suite of courses, and "ILO" emphasizes the interactive features of this kind of online learning. This is in contrast with more common types of online learning which largely mimic classroom teaching without taking advantage of the unique online environment to provide "added value," that is, anything beyond that which can be achieved in a physical classroom.

The CMU statistics course (which can be accessed at http://oli.web.cmu.edu/openlearning/) includes textual explanations of concepts and an inventory of worked examples and practice problems, some of which require the students to manipulate data for themselves using a statistical software package. Both the statistics course and other courses in the OLI suite were originally intended to be comprehensive enough to allow students to learn the material independently without the guidance of an instructor; since it was developed, however, the statistics course has been used at a variety of higher education institutions, sometimes in a hybrid mode. (Taylor Walsh describes the history of the development of this course, which was financed largely by the Hewlett Foundation over a number of years, in her 2010 book *Unlocking the Gates: How and Why Leading Universities Are Opening Up Access to Their Courses*, Princeton University Press, 2010.) Among the main strengths of the CMU statistics course is its ability to embed interactive assessments into each instructional activity, and its three key feedback loops: "system" to student, as the student answers questions; system to teacher, to inform student-instructor interactions; and system to course developer, to identify aspects of the course that can be improved. In addition to offering assessments to measure how well students understand a particular concept, the CMU course also asks students to complete self-assessments, to give the instructor and/or learning scientists a sense of how well students think they understand the concept. However, while instructors can delete and re-order modules, CMU does not offer much opportunity for customization, nor is the course adaptive in terms of redirecting students to extra practice sessions or additional reading if their incorrect answers indicate that they do not understand a concept and need more help. Thus, although the CMU statistics course is certainly impressive, we refer to it as a prototype because we believe it is an early representative of what will likely be a wave of even more sophisticated systems in the not-too-distant future.

18 See Appendix A for a description of the research protocol and incentives used on each campus.

19 The CAOS test, or Comprehensive Assessment of Outcomes in Statistics, is a 40-item multiple-choice assessment designed to measure students' statistical literacy and reasoning skills. One characteristic of the CAOS test is that (for a variety of reasons) scores do not increase by a large amount over the course of the semester. Among students in our study who took the CAOS test at both the beginning and end of the semester, the average score increase was 5 percentage points. For more information about the CAOS test, see https://app.gen.umn.edu/artist/caos.html, or delMas, Robert, Joan Garfield, Ann Ooms, and Beth Chance, "Assessing Students' Conceptual Understanding After a First Course in Statistics," 6.2 (2007): 28-58, accessed July 28, 2010, http://www.stat.auckland.ac.nz/~iase/serj/SERJ6(2)_delMas.pdf.

study participants were asked to take the CAOS test of statistical literacy again, as well as complete another questionnaire. Appendix Table A2 provides the numbers of students on each campus who were randomized into each format and who completed each data collection instrument.

Administrative data on participating and non-participating students were gathered from the participating institutions' databases. The baseline survey administered to students included questions on students' background characteristics, such as socioeconomic status, as well as their prior exposure to statistics and the reason for their interest in possibly taking the statistics course in a hybrid format. The end-of-semester survey asked questions about their experiences in the statistics course. Students in study-affiliated sections of the statistics course took a final exam that included a set of items that were identical across all the participating sections at that campus (or, in the case of the campus that had two departments participating in the study, all participating sections in that department). The scores of study participants on this common portion of the exam were provided to the research team, along with background administrative data and final course grades of all students (both participants and, for comparison purposes, nonparticipants) enrolled in the statistics course in the fall 2011 semester. All of these data are described in detail on the Ithaka S+R website, which also includes copies of the survey instruments.

Our intention was to provide a rigorous side-by-side comparison of specific learning outcomes for students in this hybrid version of the statistics course and comparable students in a traditionally-taught version of the same course. We recognize, however, that while we were reasonably successful in randomizing students between treatment and control groups (see documentation in the next section of this report), we could not randomize instructors in either group and thus could not control for differences in teacher quality.[20] This is one reason, among others, that we do not regard the research design of this project as

Our intention was to provide a rigorous side-by-side comparison of specific learning outcomes for students in this hybrid version of the statistics course and comparable students in a traditionally-taught version of the same course. However, while we were reasonably successful in randomizing students between treatment and control groups, we could not randomize instructors in either group and thus could not control for differences in teacher quality.

20 Instructor surveys reveal that, on average, the instructors in traditional format sections were much more experienced than their counterparts teaching hybrid-format sections (median years of teaching experience was 20 and 5, respectively). Moreover, almost all of the instructors in the hybrid-format sections were using the CMU online course for either the first or second time, whereas many of the instructors in the traditional-format sections had taught in this mode for years. The "experience-advantage," therefore, is clearly in favor of the teachers of the traditional-format sections. The questionnaires also revealed that a number of the instructors in hybrid-format sections began with negative perceptions of online learning. In part for these reasons, a leader of one of the sets of institutions in this study believes that results for the hybrid-format sections would be improved vis-à-vis results in the traditional-format sections if the experiment were repeated and instructors in the hybrid-format sections were better motivated and better trained. But this is, of course, a conjecture.

anything close to perfect.[21] Still, this is the first effort of which we are aware to carry out the kind of randomized study of outcomes in large introductory courses on public university campuses that we think has been needed.

One wise decision we made was to conduct spring-term pilots on as many campuses as possible in advance of the fall-term 2011 research phase of the study—when we treated outcomes as suitable for measurement. The spring-term pilots identified a number of practical aspects in which the study could be improved, and a memo on lessons learned from the spring-term pilots is included in this report as Appendix C.[22]

It remains only to add that, as Appendix C illustrates, this is *very* difficult research to do, in large part because so many details—how best to present the course, to recruit student and faculty participants, to randomize students between treatment and control groups, to collect good data including background information about the student participants, and to satisfy Institutional Review Board requirements in a timely way—need to be worked out with the day-to-day involvement of campus staff not directly responsible to us. We have great respect for other investigators who have coped with these problems, often in settings outside higher education.

Findings

The great advantage of—indeed, the main motivation for—conducting a randomized experiment is that students in the treatment and control groups are expected to have the same average characteristics, both observed and unobserved. The results in Table 2 indicate that the randomization worked properly in that traditional and hybrid-format students in fact have similar characteristics. There are a handful of small differences that are statistically significant but, in general, the differences between students taught in the traditional format and students taught in the hybrid format are not meaningful.[23]

---

21 Randomization procedures were limited by the fact that Institutional Review Board (IRB) requirements precluded randomization of students enrolled in the course without their consent. Instead, we had first to use incentives to encourage students to participate in the study, with the understanding that they would then be randomized between treatment and control groups. We were able, however, to compare the characteristics of participants and non-participants, and the two groups turned out to be very similar; see Table 3. The study is, of course, limited in that it involves only a single course, but having a common hybrid course across the six campuses (i.e. the CMU statistics course) controls for one source of variance in outcomes. We deliberately chose the CMU statistics course because we think that the greatest near-term opportunity to take advantage of interactive online technologies is in introductory-level courses that serve large student populations in fields in which there is more or less "one right answer" to most questions. Somewhat different pedagogies would be needed, we suspect, in courses that are more value-laden and dependent on discussion of various perspectives.

22 We are indebted to James Kemple, now Executive Director of the Research Alliance for New York City Public Schools, and formerly the Director of the K-12 Education Policy division at MDRC, for much useful advice. Dr. Kemple has long experience with randomized trials. Lessons learned from the pilots included how to present the project, the effective use of modest incentives for participants, and techniques that could improve randomization. We hope that others will benefit from our experience (see Appendix C) in mounting this research project.

23 A regression of format assignment on all of the variables listed in Table 2 (and institution dummies) fails to reject the null hypothesis of zero coefficients for all variables (except the institution dummies) with p=0.12. A Hotelling test fails to reject the null of no difference in means with p=0.27.

The students who participated in our study are a very diverse group. Half of the students come from families with incomes less than $50,000, and half are first-generation college students. Fewer than half are white.

In addition to testing the success of our efforts to randomize students, Table 2 also serves to describe the population of students who participated in our study. They are a very diverse group. Half of the students come from families with incomes less than $50,000 and half are first-generation college students. Fewer than half are white, and the group is about evenly divided between students with college GPAs above and below 3.0. Most students are of traditional college-going age (younger than 24), are enrolled full-time, and are in their sophomore or junior year.

These students are a diverse group, but do they resemble the entire population of students enrolled in the introductory statistics courses included in our study? Study participants were randomly assigned to a section format, but the study participants themselves are a self-selected population—because of Institutional Review Board requirements only students who agreed to be in the study were randomly assigned, and scheduling complications also limited the population of participants. Overall, 605 of the 3,046 students enrolled in these statistics courses participated in the study. An even larger sample size would have been desirable, but the logistical challenges of scheduling at least two sections (one hybrid section and one traditional section) at the same time, so as to enable students in the study to attend the statistics course regardless of their (randomized) format assignment, restricted our prospective participant pool to the limited number of "paired" time slots available. Also, as already noted, Institutional Review Boards required student consent in order for researchers to randomly assign them to the traditional or hybrid format. Not surprisingly, some students who were able to make the paired time slots elected not to participate in the study. All of these complications notwithstanding, our final sample of 605 students is by no means small—it is in fact quite large in the context of this type of research.[24]

24 Of the 46 studies examined in the Means et al. (2009) meta-analysis, only 5 had sample sizes of over 400, and of the 51 independent effect sizes the authors abstracted, 32 came from studies with fewer than 100 study participants.

TABLE 2. RANDOMIZATION OF STUDY PARTICIPANTS

	Traditional	Hybrid	Adj. Diff.	Sig?
Male	46%	39%	-7%	+
Asian	24%	23%	-1%	
Black	14%	14%	0%	
Hispanic	20%	14%	-5%	+
White	41%	46%	4%	
Other/Missing	1%	3%	2%	
Average Age	21.9	22.0	0.0	
Age <24	82%	84%	2%	
Age 24-30	14%	10%	-4%	
Age 30+	4%	5%	1%	
GPA missing	9%	5%	-4%	+
GPA <2	14%	15%	1%	
GPA 2-3	36%	38%	2%	
GPA 3+	40%	41%	1%	
Full-time	90%	90%	0%	
Freshman	11%	9%	-2%	
Sophomore	41%	46%	5%	
Junior	34%	31%	-3%	
Other/Missing	14%	13%	-1%	
Fam. income <$50k	49%	50%	2%	
Parent college grad	49%	47%	-2%	
English only lang.	65%	62%	-4%	
N	292	313		

Notes: Adjusted differences (average within-institution differences) control for institutional dummy variables. "Sig?" indicates whether the difference is statistically significant from zero at + $p<0.1$*. A regression of format assignment on all variables listed here fails to reject null of zero coefficients for all variables with* $p=.12$

TABLE 3. STUDENT CHARACTERISTICS BY STUDY PARTICIPATION

	Participant	Non-Part.	Adj. Diff.	Sig?
Male	42%	44%	-1%	
Asian	23%	17%	1%	
Black	14%	13%	0%	
Hispanic	17%	10%	3%	*
White	44%	47%	6%	*
Other/Missing	2%	13%	-10%	**
Average Age	21.9	21.6	-0.3	
Age <24	83%	81%	4%	*
Age 24-30	12%	10%	-1%	
Age 30+	5%	4%	-1%	
GPA missing	7%	13%	-7%	**
GPA <2	15%	24%	-2%	
GPA 2-3	37%	31%	3%	
GPA 3+	41%	32%	6%	**
Full-time	90%	86%	5%	**
Freshman	10%	18%	-5%	**
Sophomore	44%	40%	3%	
Junior	32%	27%	2%	
Other Year/Missing	14%	15%	0%	
Passed Course	78%	81%	-4%	*
Completed Course	84%	87%	-4%	*
Course Grade	2.37	2.36	-0.03	
N	605	2,441		

*Note: Adjusted differences (average within-institution differences) control for institutional dummy variables. "Sig?" indicates whether the difference is statistically significant from zero at ** p<0.01, * p<0.05. Students that did not complete course are assigned a course grade of zero.*

The results in Table 3 indicate that the 605 study participants, while not fully representative of all statistics students in any formal sense, have broadly similar characteristics. There are statistically significant differences between study participants and non-participants on several characteristics, but most of the differences are small in magnitude. For example, participants are more likely to be enrolled full-time, but only by a margin of 90 versus 86 percent. Course outcomes are also broadly similar, with participants earning similar grades and being only slightly less likely to complete and pass the course as compared to non-participants.

Our analysis of the data is straightforward; we compare the outcomes of students randomly assigned to the traditional format to the outcomes of students randomly assigned to the hybrid format. In a small number of cases—4 percent of the 605 students in the study—participants attended a different format section than the one to which they were randomly assigned. In order to preserve the randomization procedure, we associated students with the section type to which they were randomly assigned. This is sometimes called an "intent to treat" analysis. Under certain assumptions, the effect of actually taking the course in the hybrid format (as opposed to just being randomly assigned to do so) can be calculated by increasing our estimates by 4 percent.[25] This is sometimes called the "treatment on the treated" estimate, which in our study is very similar to the "intent to treat" estimate because most students took the course in the format to which they were randomly assigned.

How did learning outcomes compare across the treatment and control groups? We first examine the impact of assignment to the hybrid format, relative to the traditional format, in terms of the rate at which students completed and passed the course, their performance on a standardized test of statistics (the CAOS test), and their score on a set of final exam questions that were the same in the two formats.[26] Our main results are summarized in Figure 1 (page 19), and the regression results are reported in Appendix Table A3.[27] We find no statistically significant differences in learning outcomes between students in the traditional – and hybrid-format sections. Hybrid-format students did perform slightly better than traditional-format students on three outcomes, achieving pass rates that

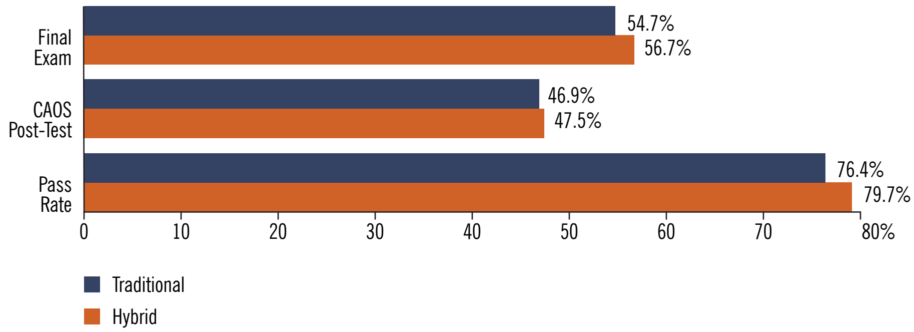
25 The key assumption is that being randomly assigned to hybrid or traditional did not have an effect on student outcomes independent of its effect on the format in which students were enrolled. This assumption would be violated if, for example, students hoped for a certain outcome of the random assignment and were disappointed when they did not get their preferred assignment, which in turn caused them to do worse in the course.

26 All of our results control for course-specific dummy variables, since students were randomized within courses; these variables also control for unobserved student characteristics that are constant within institutions. However, we obtain similar results when we do not control for institution dummies, as would be expected given that the probability of being assigned to the hybrid section was constant across courses (50%).

27 Note that the pass rate in Figure 1 and Appendix Table A3 cannot be used to calculate the percentage of students who failed the course because the non-passing group includes students who never enrolled or withdrew from the course without receiving a grade.

wereabout three percentage points higher, CAOS scores about one percentage point higher, and final exam scores two percentage points higher—but none of these differences passes traditional tests of statistical significance.[28]

Figure 1. Effect of Hybrid Format on Student Learning Outcomes



Notes: None of the traditional-hybrid differences above were statistically significant at the 10% level. See Appendix Table A3 for more information about the results depicted here.

In other words, we can be quite confident that the 'average' effects were in fact close to zero. We also find that the same basic results hold for subgroups, and that distributions of key outcomes are very similar for both the treatment and control group students.

It is important to report that these differences (or rather, the lack of statistically significant differences) are fairly precisely estimated—see both the actual coefficients and the small standard errors of the effect estimates reported in Appendix Table A3.[29] In other words, we can be quite confident that the "average" effects were in fact close to zero. As we explain shortly, we also find that the same basic results hold for subgroups, and that distributions of key outcomes are very similar for both the treatment and control group students. One commentator, Michael S. McPherson, president of the Spencer Foundation, observed that what we have here are "quite precisely estimated zeros."

That is, if there had in fact been pronounced differences in outcomes between traditional-format and hybrid-format groups, it is highly likely that we would have found them.[30] Our findings are strikingly different in this consequential respect from a hypothetical finding of "no significant difference" which resulted from a coefficient of some magnitude accompanied by a very large standard error or by big differences in the distributions of outcomes. A hypothetical finding of this kind would mean, in effect, that we don't know much: that the "true" results could be almost anywhere.

28 The effect on CAOS test scores in standard deviation units (using the distribution of the control group) was 0.05. We also examined performance using item-level CAOS post-test data. Specifically, we grouped the 40 items into the 20 items on which delMas et al.'s (2007) national sample of students exhibited significant growth (over the course of a semester) and the remaining 20 items. We found similar hybrid-format effects for the two groups of items.

29 We cluster standard errors by section in order to capture section-specific shocks to student outcomes (such as the quality of the instructor). Students who were randomly assigned but never enrolled in the course are grouped as a "section" within each course for the purpose of computing clustered standard errors.

30 Some degree of caution is warranted in interpreting the results for the CAOS post-test because the average student's CAOS score only increased by five percentage points over the course of the semester (among students who took both the pre-test and the post-test). This may have resulted in part from some students not taking the CAOS test seriously because, in most cases, it was not part of their grade. However, it is reassuring that the results for the CAOS test are consistent with results for pass rates and final exam scores.

Results broken down by individual institution do not reveal any noteworthy patterns.

These findings control for student characteristics, including race/ethnicity, gender, age, full-time versus part-time enrollment status, class year in college, parental education, language spoken at home, and family income. These controls are not strictly necessary since students were randomly assigned to section format, but we include them in order to increase the precision of our results and to control for any remaining imbalance in observable characteristics. However, we obtain nearly identical results when we do not include these control variables—just as we would expect given the apparent success of our random assignment procedure.

Our results are also robust to a variety of alternative methodologies used to analyze the experimental data. These results are reported in Appendix Table A4, and one is worth highlighting. A limitation of our main results for CAOS post-test and final exam scores is that we only observe these outcomes for students who completed the course and took these exams. This is unlikely to be a significant limitation given that we do not find any significant effects of section format on course completion and pass rates. But as an additional check, we assigned students for whom we did not observe a CAOS post-test score their score on the CAOS pre-test—in other words, we assumed that their score did not change over the course of the semester. Students who did not take either the pre-test or the post-test were assigned the average pre-test score at their institution. The resulting set of real and imputed post-test scores yielded very similar results to those obtained using only the real data.

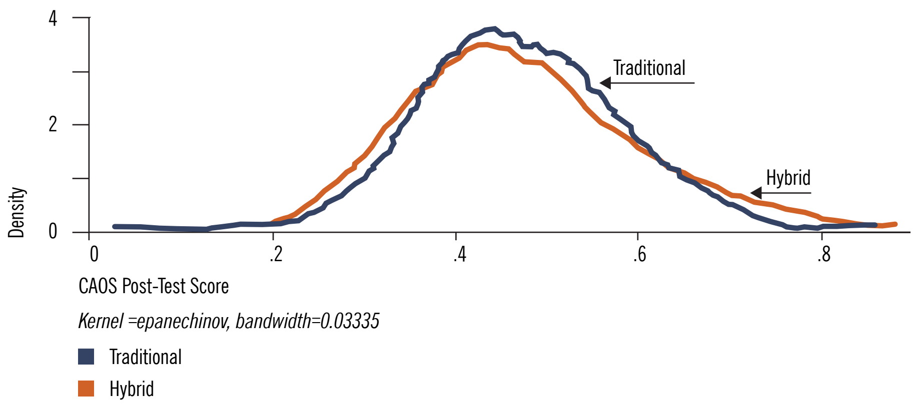
The lack of differences in mean outcomes between formats could mask differences in the distribution of outcomes.[31] Figure 2 (page 21) shows that this is not the case for CAOS post-test scores. The distributions of scores for traditional and hybrid format students are largely similar, although scores are slightly more spread-out for hybrid-format students. We obtain a similar finding for final exam scores (not shown).[32] (This kind of comparison of distributions is not possible for pass rates, which only take on a value of 0 or 1 for an individual student.)

Results broken down by individual institution (Appendix Table A5) do not reveal any noteworthy patterns. These results are much noisier because they are based on smaller numbers of students, but they do not indicate that the hybrid format was particularly effective or ineffective at any individual institution—with the possible exception of Institution F, where coefficients are positive across all four outcomes, although only statistically significant in the case of one outcome.

31 We are indebted to Stephen Stigler, a professor at the University of Chicago and a member of the ITHAKA board, for emphasizing to us the importance of considering this question.

32 In general, results that use final exam scores should be interpreted cautiously given limitations in these exams and their implementation. Some institutions included only a handful of questions that were common across the sections of the course (and we only use data from the common questions). At one institution, common questions were administered to some students after the end of the semester because the actual final exam only included common questions in two out of six sections. At another institution, final exam data were not available for the students of two instructors (covering three out of six traditional-format sections). Excluding either or both of these institutions produces similar results, but it should be noted that the effect on final exams in standard deviation units is substantial in size (0.19; see Appendix Table A4) and imprecisely estimated.

Figure 2. Distributions of CAOS Post-Test Scores



We also calculated results separately for subgroups of students defined in terms of various characteristics, including race/ethnicity, gender, parental education, primary language spoken, CAOS pre-test score, hours worked for pay, and college GPA. We did not find any consistent evidence that the hybrid-format effect varied by any of these characteristics (Appendix Table A6).[33] There were no groups of students that benefited from or were harmed by the hybrid format consistently across multiple learning outcomes.

Our main results provide compelling evidence that, on average, students learned just as much in the hybrid format as they would have had they instead taken the course in the traditional format—with "learning" measured in traditional ways, in terms of course completion, course grades, and performance on a national test of statistical literacy. This seemingly bland result is in fact very important, in light of perhaps the most common reason given by faculty and deans for resisting the use of ILO-type instruction: "We worry that basic student learning outcomes (pass rates and mastery of content) will be hurt, and we won't expose our students to this risk." The research reported here suggests strongly that such worries are not well founded.

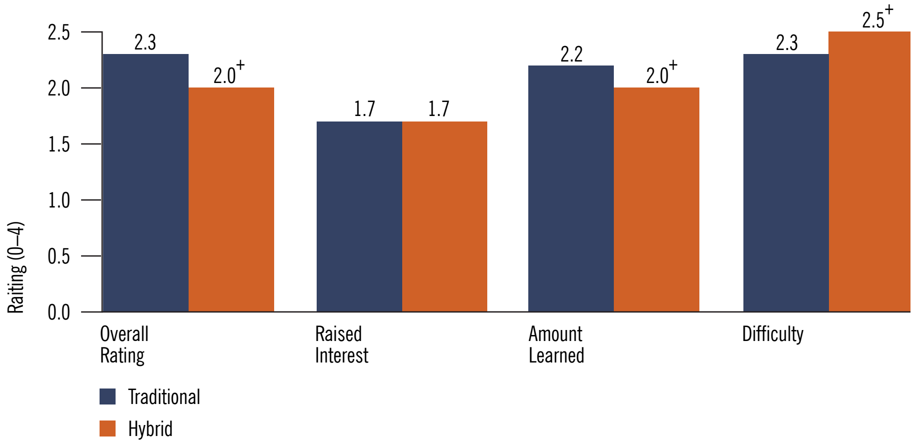
We also examined how much students liked the hybrid format of the course relative to traditional format students (Figure 3 [page 22] and Appendix Table A7). We found that students gave the hybrid format a modestly lower overall rating than the one given by students taking the course in traditional format (the rating was about 11 percent lower). By similar margins, hybrid students reported feeling

33 The one exception is our finding that completion and pass rates were significantly higher in the hybrid course for students with family incomes of at least $50,000 per year, but not for students with family incomes of less than $50,000 per year. However, we hesitate to attach much significance to this result given that we do not find such a pattern for our other measure of socioeconomic status (parental education) or for measures of academic preparation. This finding could be the result of random noise in the coefficients (especially given the large number of coefficients reported in Appendix Table A6).

Our results indicate that hybrid-format students took about one-quarter less time to achieve essentially the same learning outcomes as traditional-format students.

that they learned less and that they found the course more difficult.[34] These three differences, though modest in size, were statistically significant at the 10 percent level. But there were no statistically significant differences in students' reports of how much the course raised their interest in the subject matter.

Figure 3. Effect of Hybrid Format on Student Evaluations of Course



Notes: + indicates traditional-hybrid difference is statistically significant at the 10% level. See Appendix Table A7 for more information about the results depicted here.

A leader of one of the universities that actively participated in our study opined that a defect of the CMU prototype course is that it has no "addictive" or "Disney-like" appeal; it was, as this person put it, "designed by cognitive scientists" (no offense intended!). In contrast, some students in the traditional format may have been treated to an occasional colorful story, personal recollections of the instructor, or other "treatments" sometimes used by faculty to improve students' opinions of their course.[35]

We also asked students how many hours per week they spent *outside of class* working on the statistics class. Hybrid-format students reported spending 0.3 hour more each week, on average, than traditional-format students. This differ-

34 Students' responses to the open-ended questions on the end-of-semester surveys indicate that many students in the hybrid format would have liked more face-to-face time with the instructor than one hour each week; others felt that the instructor could have better used the face-to-face time to make the weekly sessions more structured and/or helpful in explaining the material and going over concepts students did not understand. A number of students in the hybrid course also indicated they would have benefited from more practice problems or examples, and many were frustrated by the difficulty of checkpoint assessments in the course and by problems they encountered using the statistical software packages to complete assignments.

35 The question of what is really going on here—with no differences in learning outcomes as measured conventionally combined with a (to be sure, small) difference in qualitative assessments—relates to a larger literature on measured learning outcomes versus more "subjective" measures of student satisfaction with online or hybrid courses relative to their satisfaction with face-to-face courses. Studies (some more rigorous than others) pertaining to the latter topic abound; a few examples include: Hannay, Maureen, and Tracy Newvine. "Perceptions of Distance Learning: A Comparison of Online and Traditional Learning." 2.1 (2006): 1-11. Accessed April 24, 2012. http://jolt.merlot.org/documents/MS05011.pdf; Horspool, Agi, and Carsten Lange. "Applying the Scholarship of Teaching and Learning: Student Perceptions, Behaviours and Success Online and Face-to-Face." 37.1 (2011): 73-88. Accessed April 24, 2012. http://www.tandfonline.com/doi/abs/10.1080/02602938.2010.496532; and Meyer, Katrina A. "Student Perceptions of Face-to-Face and Online Discussions: The Advantage Goes To ...." .4: 53-69. Accessed April 24, 2012. http://sloanconsortium.org/jaln/v11n4/student-perceptions-face-face-and-online-discussions-advantage-goes.

ence, which is not statistically significant, implies that, in a course where a traditional section meets for 3 hours each week and a hybrid section meets for 1 hour, the average hybrid-format student would spend 1.7 less hours each week in *total time* devoted to the course, a difference of about 25 percent. This result is consistent with other evidence that ILO-type formats do succeed in achieving the same learning outcomes as traditional-format instruction in less time—which potentially has important implications for scheduling and the rate of course completion.[36]

In sum, our results indicate that hybrid-format students took about one-quarter less time to achieve essentially the same learning outcomes as traditional-format students. The three main limitations of this analysis are: (1) we were not able to randomly assign instructors to section formats—which would have been difficult, if not impossible, to do, especially in a small scale study, or to control for differences in how traditional-format sections were taught;[37] (2) the limitations of the CMU prototype of an ILO course—no customization and no "addictive" features; and (3) our inability to report results for community colleges. Despite these limitations, these results reflect a serious, rigorous, assessment of the relative efficacy of technology-enhanced learning (ILO-style hybrid instruction) compared to the traditional mode of instruction. They are, we believe, the best evidence available to date on an important set of questions. There is, without doubt, much more research that can and should be carried out but, at the minimum, this study supports a "no-harm-done" conclusion regarding at least one current prototype of an ILO system.

36 The authors of this paper, interested to see whether the hybrid course might enable students to learn the material in the statistics course in a shorter period of time, conducted a separate, quasi-experimental study in the summer of 2011, involving one of the campuses used in our larger fall 2011 study. This summer "probe" occurred over two shortened (five – or six-week) summer sessions; while we collected a large amount of data and used numerous controls in our analysis, we did not randomly assign students to the hybrid or the face-to-face format so that we could obtain larger sample sizes (from what was already a much smaller pool of students taking the course during the summer). The results of the summer study revealed no statistically significant differences in the percentage of students who passed the course, in final exam grades, or in the end-of-semester CAOS test results, between the students who took the course in a hybrid format and students who took the course in a face-to-face format. A separate study by Marsha Lovett, associate director of the Eberly Center for Teaching Excellence, and her colleagues at CMU produced similar findings. In that study, which was conducted at CMU, the performance of students who were randomly assigned to the face-to-face format of a statistics course, which met four times a week for 50 minutes each time, was compared with the performance of students who were randomly assigned to a hybrid format, which met twice a week for 50 minutes each time. The researchers found little difference in the amount of time students reported spending on the course outside of class each week (about 2.8 hours for the hybrid-format students, compared with about 2.7 hours per week for face-to-face students). In addition, students also were able to learn in eight weeks the same amount of material that students would ordinarily take 15 weeks in a face-to-face format to learn. In this respect, the use of the CMU course could be said to increase learning efficiency. (For more information about this study, see Lovett, Meyer, and Thille's 2008 article in the , entitled "The Open Learning Initiative: Measuring the effectiveness of the OLI statistics course in accelerating student learning," available online at http://jime.open.ac.uk/2008/14.)

37 One commentator, Michael McPherson, noted that right now, quite apart from any use of online technologies, very different instructional methods are used to teach introductory statistics courses even within the same university—never mind across universities. But little effort appears to be made to compare learning outcomes and cost effectiveness across different approaches—a point Derek Bok, former president of Harvard who continues to write on this subject, keeps emphasizing.

It is an open question whether subsequent progress in improving hybrid ILO courses of this kind—in particular, efforts to achieve greater customization opportunities for faculty,[38] and to make it more "fun" (more addictive) for both students and faculty to use the system—will lead to more positive conclusions concerning learning effectiveness. We just don't know. Digital learning is still in early days, and it is entirely possible that future versions of hybrid courses will not just maintain the same learning outcomes but increase student learning relative to the status quo. But we do know enough now to justify careful exploration of potential cost savings.

Productivity—measured as outputs divided by inputs—has been increased by the use of technology in other sectors of the economy, often resulting in increased output. Our multi-campus study of learning outcomes in undergraduate education has shown that a leading prototype hybrid learning system did not lead to a statistically significant increase in outputs (student learning), but could potentially increase productivity nonetheless by using fewer inputs—thereby achieving cost savings. Until now, technology-induced productivity gains in higher education have been taken mainly in the form of increased output—more and faster research, and so on; the time may be at hand when cost savings should be sought and emphasized.[39]

We conceptualize the research question here not as "how much will institutions save right now by shifting to hybrid learning?" but rather as "under what assumptions will cost savings be realized, over time, by shifting to a hybrid format, and how large are those savings likely to be?

The cost blade of the scissors is at least as challenging to study as the learning blade. This may be one reason why so few studies have paid much attention to costs. (Carol Twigg's work with the National Center for Academic Transformation project is a notable exception.[40]) At first blush, it would seem to be straightforward to compare the side-by-side costs of the hybrid-online version of the statistics course and the traditional version. Not so. Our early efforts to do just that were unsuccessful. The big problem, we learned, is that contemporaneous comparisons can be near-useless in projecting steady-state costs because the costs of doing almost anything for the first time are very different from the costs of doing the same thing numerous times. That admonition is especially true in the case of online learning. The cost implications of some educational interventions can be measured immediately and with relatively little difficulty. For example, the higher cost associated with a decrease in the size of a class is simply the cost of the additional instructors and space required to accommodate a larger number of classes with fewer students in each one. This cost will be more or less the same in the first year the intervention is implemented as in the tenth year.

In the case of hybrid learning, however, there are substantial start-up costs that have to be considered in the short run but are likely to decrease over time, thereby making short-term costs significantly greater than long-term costs. For example, the development of sophisticated hybrid courses will be a costly effort that would only be a sensible investment if the start-up costs were either paid for by others (foundations and governments) or shared by many institutions and

38 The pervasive desire for more ability to customize is a key finding of "Barriers" by Bacow et al. (see pages 21-22 in particular).

39 This is another conclusion of the "Barriers" study by Bacow et al.; see pages 22 and 24-25 in particular.

40 See the National Center for Academic Transformation website at www.thencat.org.

amortized over time. There are also transition costs entailed in moving from the traditional, mostly face-to-face model (that may, however, employ some elements of simple online models, such as video-taped lectures or homework assignments online) to a hybrid model that takes advantage of more sophisticated ILO systems employing machine-guided instruction, cognitive tutors, embedded feedback loops, and some forms of automated grading. Instructors need to be trained to take full advantage of such systems. There may also be contractual limits on section size that were designed with the traditional model in mind but that do not make sense for a hybrid model. It is possible that these constraints can be changed in a next round of contract negotiations, but that too will take time.

To overcome (or avoid!) these problems, we think there is much to be said for carrying out simulated "cost probes," and we made a very rough attempt to do just this on three of the campuses included in the "learning outcomes" part of the study. We conceptualize the research question here not as "how much will institutions save right now by shifting to hybrid learning?" but rather as "under what assumptions will cost savings be realized, over time, by shifting to a hybrid format, and how large are those savings likely to be?" Our basic approach was to start by looking in as much detail as possible at the actual costs of teaching a basic course in traditional format (usually, but not always, the statistics course) in a base year. Then, we worked with leaders on these campuses to simulate the prospective, steady-state costs of a hybrid-online version of the same course, looking three to five years into the future. These exploratory simulations were based on explicit assumptions, especially about staffing, that were incorporated into spreadsheets—which in turn allowed us to see how sensitive our results were to variations in key assumptions. We focused heavily on personnel costs, because of both their importance and our ability to examine them with some precision. Other costs, including space costs, were also considered. We hoped that the simulations would, at the minimum, give us at least a rough sense of the potential impact on costs of introducing hybrid learning and, more specifically, show us how much "room" there would be for institutions to share cost savings with faculty and students on a continuing basis.

Our simulations illustrate that hybrid learning offers opportunities for significant savings in compensation costs, but that the degree of cost reduction depends on the exact model of hybrid learning used.

We focus on instructor compensation because these costs comprise a substantial portion of the recurring cost of teaching and are the most straightforward to measure. Space costs are also an important category of costs that are likely to be reduced by shifting to a hybrid learning model (the most important category in some situations), but such costs are more difficult to measure accurately at the level of an individual course. A hybrid model also affords both faculty and students significantly greater scheduling flexibility, a potentially very important benefit that will not be captured by our simulations. On the other hand, there are also other types of costs that we do not consider here, such as increases in information technology (IT) support costs associated with moving a significant share of learning activities online. Such added costs can be far from trivial.

We did exploratory simulations for two types of traditional teaching models: (1) a model in which students are taught in sections of roughly 40 students per section; and (2) a model in which all students attend a common lecture and are then assigned to small discussion sections led by teaching assistants. We compare the current costs of each of these traditional teaching models to simulated costs of a hybrid model in which more instruction is delivered online, students attend weekly face-to-face sessions with part-time instructors, and the course is overseen by a tenure-track professor (with administrative responsibilities delegated to a part-time instructor).

We have decided, however, not to present the actual calculations and results of these simulations in the body of this report. They are too speculative and subject to considerable variation depending on how a particular campus wanted to organize its teaching. The danger of "specious precision" is great, and it would be wrong to attach much importance to particular numbers. Suffice it to say that the crude models we employed suggest savings in compensation costs ranging from 36 percent to 57 percent in the all-section model, and 19 percent in the lecture-section model. Appendix B presents these results and many more calculations and figures showing how sensitive potential savings are as we vary assumptions about section sizes and compensation.

These simulations illustrate that hybrid learning offers opportunities for significant savings in compensation costs, but that the degree of cost reduction depends (of course) on the exact model of hybrid learning used—especially the rate at which instructors are compensated and section size. A large share of cost savings is derived from shifting away from time spent by expensive professors toward both machine-guided instruction that saves on staffing costs overall and toward time spent by less expensive staff in Q&A settings. Of course, tenured professors cannot be laid off in order to realize these savings and, in any case, "force reductions" are not required to save significant amounts of money. Institutions that face pressures to expand enrollment are in an especially good position to realize savings by shifting the mix of teaching time. When more students are to be taught, hybrid models make this possible without increasing the demands made on tenured faculty. Recruitment costs may thereby be reduced along with compensation costs per student, and debates over maintaining commitments to existing faculty are avoided. Over time, certainly, staff size can be altered through attrition. Also, the time of professors can be reallocated toward smaller, more advanced classes—which many prefer to teach (such reallocations may not save the institution money, though they may improve the overall educational experience of many students).

In these simulations, we have assumed that the number of students in the course will remain constant. However, as already suggested, many institutions face increasing demand for places in their classes. The hybrid learning model is very attractive in such circumstances for two primary reasons: (a) less space is needed in general; and (b) hybrid courses provide both students and teachers with greater scheduling flexibility. Increased enrollment can also lead to increased compensation cost savings (per student) because the fixed costs of the professor in charge of the course, and an administrative coordinator, would be spread over a larger number of students. For the same reason, the largest savings will be realized in courses with the largest enrollment, all else equal.

Our simulation approach underestimates substantially the savings from moving toward a hybrid model in many settings because we do not account for space costs. We are reluctant to put a dollar figure on space costs because capital costs are difficult to apportion accurately down to the course level. However, it is more straightforward to calculate the percentage change in the need for classroom space that would result from shifting toward a hybrid model. The hybrid course meets for one hour each week, whereas the traditional course typically meets for three or four hours each week. Consequently, the hybrid course requires 67 percent to 75 percent less classroom use than the traditional course, assuming that the course is taught in sections, that section size is held constant, and that the hybrid course does not have additional space requirements of its own, such as additional computer labs.

In the short run, institutions cannot sell or demolish their buildings. However, in the long run, using hybrid models for some large introductory courses would allow institutions to expand enrollment without a commensurate increase in space costs—a major cost savings (cost avoidance) relative to what institutions would have had to spend had they stayed with a traditional model of instruction. An important point here is that the hybrid model need not just "save money"—it can also support an increase in access to higher education. It serves the access goal both by making it more affordable for the institution to enroll more students and by accommodating more students because of greater scheduling flexibility, which is especially important for students with complicated lives who have to balance family responsibilities and work with course completion, as well as for students who may live a distance from campus.[41]

To repeat, we regard this highly speculative cost simulation effort as primarily illustrative of an approach that we believe has merit. It is no surprise that under a plausible set of assumptions ILO systems have the potential to save staffing costs and classroom space, and to increase scheduling flexibility. To go beyond that obvious statement requires, at the minimum, detailed knowledge of local campus situations and realistic assessments of what is feasible.

Summary Observations

In the case of a topic as "active" as online learning, where new articles appear every day and millions of dollars are being invested by a wide variety of entities, we should perhaps expect that there will be inflated claims of spectacular successes. The findings in this study warn against "too much hype." To the best of our knowledge, there is no compelling evidence that online learning systems available today—not even highly interactive systems, of which there are very few—can in fact deliver improved educational outcomes across the board, at scale, on campuses other than the one where the system was born, and on a sustainable basis. This is not to deny, however, that these systems have great potential. We believe that they do, and that vigorous efforts should be made to aggressively explore uses of both the relatively simple systems that are proliferating all around us, often to good effect, and more sophisticated systems that are still in their infancy. There is every reason to expect these systems to improve over time, perhaps dramatically, and thus it is not foolish to believe that learning outcomes will also improve.

41 The MOOCs (Massive Open Online Courses), like edX and Udacity, are another example of a technologically-driven effort to give many more students, of all kinds, access to low-cost instruction of high quality. It remains to be seen, however, whether the certificates and badges that MOOCs propose to confer will be accepted as credits toward degrees by mainline universities—and how much this will matter to students themselves, as well as to employers and others who want to assess learning outcomes.

The barriers to adoption of even the simpler systems, let alone those of more sophisticated systems that are truly interactive, are detailed in the companion Ithaka S+R report by Lawrence S. Bacow et al. that has been cited previously.[42] There is no need to summarize those findings here. It is sufficient to re-emphasize the need to:

a. work closely with faculty, who understandably will want to put at least something of their own stamp on all such courses;

b. confront directly and imaginatively worries about loss of jobs;

c. encourage serious research by trusted third parties on evidence of learning outcomes;[43] and

d. recognize, and even embrace, the need to use such technologies to increase productivity and lower instructional costs without sacrificing learning outcomes.

We are persuaded that well-designed interactive systems have the potential to achieve at least equivalent educational outcomes while opening up the possibility of saving significant resources which could then be redeployed more productively.

The research reported here demonstrates the potential of truly interactive learning systems that use machine-guided protocols (what we have been calling "ILO") to provide some forms of instruction, in properly chosen courses, in appropriate settings. Our findings demonstrate that such an approach need not negatively impact learning outcomes—and conceivably could, in the future, improve them as these systems become ever more sophisticated and user-friendly. It is also entirely possible that by (potentially) saving significant amounts of resources, such systems can lead to more, not less, opportunity for students to benefit from exposure to modes of instruction such as directed study—if scarce faculty time can be beneficially redeployed. But none of this will be easy.

In the spirit of "thinking out loud," here are some thoughts as to what is required if we are to take full advantage of the potential of ILO offerings. These observations, we should note, are based less on the specific findings in this study than on conversations with a wide range of individuals concerned with the future of online education—and, for that matter, with the future of higher education.[44]

- First, a positive but realistic mindset on everyone's part is essential. The challenges here are at least as much organizational as they are technical. Time and patience (a reasonably long time-horizon) will be required. The greater use of technology in teaching could benefit everyone, and the implications of various approaches for models of shared governance deserve a great deal of thought.

42 See "Barriers to Adoption of Online Learning Systems in U.S. Higher Education" by Bacow et al.

43 One thing we learned from interactions with campuses is that often there is no agreement, even at the course level, as to what students should learn. Different final exams were used in various sections of the same course. As ILO-type instruction is tried more widely, it will become even more important to get agreement on learning outcomes and on how they are to be measured. Another potential benefit, then, of trying out such approaches is that the use of such systems may encourage (if not require) more direct consideration of what students are expected to learn.

44 These same conversations formed the basis of Bacow et al.'s "Barriers" study.

- Second, a major conclusion of the Bacow report is that "perhaps the largest obstacle to widespread adoption of ILO-style courses" is the lack at the present time of a "sustainable platform that allows interested faculty either to create a fully interactive, machine-guided learning environment or to customize a course that has been created by someone else (and thus claim it as their own)."

- Third, a system-wide approach will be needed if a sophisticated customizable platform is to be developed, made widely available, maintained, and sustained in a cost-effective manner. It is unrealistic to expect individual institutions to make the up-front investments needed to create such a platform, to extend its use broadly, and to sustain it. It is also widely recognized that collaborative efforts among institutions are difficult to organize, especially when much nimbleness is needed. In all likelihood, either major foundation or governmental investments will be required to launch such a project. It is conceivable that initiatives such as edX (at MIT and Harvard) could evolve to serve the needs of large public universities, but that remains to be seen. Ambitious efforts to develop a common platform could of course fail, and clear expectations should be set before they are tried. Still, as the saying goes, "nothing ventured, nothing gained."

- Fourth, as new ILO courses are developed in different fields (perhaps based on a new platform, as suggested above), it will be important to test them out rigorously, to see how cost-effective they are in at least sustaining and possibly improving learning outcomes for various student populations in a variety of settings. Such rigorous testing should be carried out in large public university systems which may be willing to pilot such courses. Hard evidence will be needed to persuade other institutions, and especially leading institutions, to try out such approaches.

- Fifth and finally, it is hard to exaggerate the importance of confronting directly the cost problems facing all of higher education and especially the public sector. As we argue in the introduction to this report, the public is losing confidence in the ability of the higher education sector to control cost increases. All of higher education has a stake in addressing this problem, including the elite institutions that are under less immediate pressure than others to alter their teaching methods. ILO systems can be helpful not only in curbing cost increases (including the costs of building new space), but also in improving retention rates, educating students who are place-bound, and increasing the throughput of higher education in cost-effective ways.

We do not mean to suggest—because we do not believe—that ILO systems are some kind of panacea for this country's deep-seated educational problems, which are rooted in fiscal dilemmas and changing national priorities as well as historical practices. Many claims about "online learning" (especially about simpler variants in their present state of development) are likely to be exaggerated. But it is important not to go to the other extreme and accept equally unfounded assertions that adoption of online systems invariably leads to inferior learning outcomes and puts students at risk. We are persuaded that well-designed interactive systems have the potential to achieve at least equivalent educational outcomes while opening up the possibility of saving significant resources which could then be redeployed more productively. Emerging interactive online systems represent one opportunity to "bend cost curves" in educationally responsible ways—and, at the minimum, to demonstrate a willingness to confront today's problems in new ways.

Acknowledgements

We wish to thank the following individuals for their help with this research and report:

From the City University of New York

Paul Bachler, Sophia Barrett, Sadie Bragg, John Choonoo, Jane Delgado, Linda Friedman, Annie Han, Alexandra Logue, Joan Lucariello, James McCarthy, Ellen Smiley, Zun Tang, and Mari Watanabe;

From the State University of New York

Joel Bloom, Carmine Desanto, Carol Farber, Fred Hildebrand, David Lavallee, Karin Reinhold, William Thistleton, Bruce Szelest, and Nancy Zimpher;

From the University System of Maryland and Montgomery College

Laura Bageant, Deneen Blair, Greg Chasson, Terry Cooney, Linda Cooper, David Earnest, Bonnie Kegan, William E. Kirwan, William Lacourse, Margaret Latimer, Nagaraj Neerchal, and Nancy Shapiro;

From Carnegie Mellon University

Marsha Lovett, Oded Meyer, John Rinderle, and Candace Thille.

In addition, we wish to thank all of the instructors involved in our project at the institutions listed above, without whom this research would not have been possible.

Appendices

Appendix A: Additional Tables

TABLE A1: SUMMARY OF RESEARCH PROTOCOLS*

Institution	Incentives	Recruitment, Registration, and Randomization
Institution A	• *Control group*: $50 American Express gift card • *Treatment group*: No textbook purchase required	• Students registered for traditional sections as normal. On the first day of class, students were given a presentation and invited to participate. Students could sign consent forms in class or return them by the next day. After all consent forms were received, participants were randomized and then informed of their format assignments by email. • *Control group*: 2 sections of 181 and 227 students (mix of participants/non-participants) • *Treatment group*: 2 sections of 18 and 28 students (participants only)
Institution B	• *Control group*: free e-textbook • *Treatment group*: No textbook purchase required	• Students were recruited both before advance registration in spring, and during the first few weeks of the fall semester, via information sessions, website, flyers, and ads in student newspaper. 3 control/treatment section pairs in 3 time slots reserved for participants; students registering for those sections received email with link to consent form; students who did not complete the consent form within 7 days automatically dropped. Participants were randomized about a week before classes began, and then notified by email. • *Control group*: 3 sections, ranging from 30 to 45 students (participants only) • *Treatment group*: 3 sections, ranging from 18 to 36 students (participants only)
Institution C	• *Control group*: $50 Amazon gift card and priority registration for spring 2013 • *Treatment group*: $50 Amazon gift card and priority registration for spring 2013	• Students were recruited via email from department chair before advance registration. All students were required to register for the course in person at computer lab, where they were informed about the study and sent email invitation with link to consent form. Students checked email on the spot and those who agreed to participate were randomized and immediately informed of their assignment. They could then choose what time slot to register for within their assigned format. • *Control group*: 6 sections from 19 to 26 students (mix of participants/non-participants) • *Treatment group*: 4 sections from 16 to 23 students (mix of participants/non-participants)
Institution D	• *Control group:* $50 book store gift card • *Treatment group:* No textbook purchase required	• Students registered for traditional sections as normal. On the first day of class, students were given a presentation and invited to participate. Students could sign consent forms in class or return them by the next day. After all consent forms were received, participants were randomized and then informed of their format assignments by email. • *Control group*: 2 sections of 29 and 37 students (mix of participants/non-participants) • *Treatment group*: 2 sections of 3 and 5 students (participants only)
Institution E	• Control group *(Departments 1 & 2)*: $50 in credit to university account • Treatment group *(Departments 1 & 2)*: no textbook purchase required**	• Students registered for traditional sections as normal. On the first day of class, students were given a presentation and invited to participate. Students could sign consent forms in class or return them by the next day. After all consent forms received, participants randomized and then informed of their format assignments by email. • *Control group (department 1)*: 3 sections from 14 to 30 students (mix of participants/non-participants) • *Treatment group (department 1)*: 1 section with 15 students (participants only) • *Control group (department 2)*: 2 sections of 24 and 73 students (mix of participants/non-participants) • *Treatment group (department 2)*: 2 sections of 11 and 14 students (participants only)
Institution F	• *Control group*: $25 gift card and free textbook • *Treatment group*: $25 gift card; no textbook purchase required	• Students were recruited via flyers during freshman orientation. All students required to register in person in department 1, where they were informed about the study and directed to an online consent form. Students who consented were randomly assigned and immediately informed of their format assignment. • *Control group*: 1 section with 90 students (mix of participants/non-participants) • *Treatment group*: 1 section with 99 students (mix of participants/non-participants)

For additional details about the protocol implemented at each campus, see http://www.sr.ithaka.org.

***At Institution E, students in the treatment group who had Macintosh computers were also given software that allowed them to run the Windows operating system in order to complete course assignments using the statistical package Minitab.*

TABLE A2: RANDOM ASSIGNMENTS OF STUDENTS AND COMPLETION OF DATA COLLECTION INSTRUMENTS

(T= Treatment, C=Control)

Institution	Random Assignment		Enrollments in Study-Affiliated Sections at Start of Term		Switched Formats		Completed Course		Completed Survey Instrument				
	T	C	T	C	T to C	C to T	T	C	Baseline Survey	Pre CAOS test	Post CAOS test	Common Final Exam Questions	End of Semester Survey
Institution A*	52	45	52	45	2	0	42	36	90	83	70	70	64
Institution B	117	112	111	108	3	0	100	91	229	209	189	63	188
Institution C	47	45	40	40	2	4	35	37	92	69	50	52	49
Institution D*	9	7	9	7	1	0	8	6	15	15	13	12	13
Institution E, Department 1*	15	16	15	16	0	0	12	11	31	31	22	23	22
Institution E, Department 2*	26	24	26	24	2	0	22	25	49	50	44	45	45
Institution F	47	43	48	30	3	9	48	30	90	74	70	76	60

**The "Enrollments in Study-Affiliated Sections" columns indicate how many participants were enrolled in that section and format* immediately after *randomization had occurred. (At these campuses, randomization took place within the first week of class rather than before the start of the semester.) This does not take into account students who switched formats midway through the term. Thus, the "Random Assignment" and "Enrollments in Study-Affiliated Sections" figures will be the same at these campuses.*

TABLE A3. HYBRID EFFECTS ON LEARNING OUTCOMES

	Without Controls			
	Complete	**Pass**	**CAOS Post**	**Final Exam**
Hybrid	0.05	0.04	0.00	0.03
	[0.03]	[0.04]	[0.01]	[0.03]
Observations	605	605	458	431
	With Controls			
	Complete	Pass	CAOS Post	Final Exam
Hybrid	0.04	0.03	0.01	0.02
	[0.03]	[0.04]	[0.01]	[0.02]
Observations	605	605	458	431
Control Mean	0.82	0.76	0.47	0.55
Control SD	–	–	0.11	0.22

*Notes: ** p<0.01, * p<0.05, + p<0.1. Results for Complete and Pass rates are marginal effects from probit regression. All results control for institution dummies. The results in the bottom panel also control for student race/ethnicity, gender, age, full-time status, year in college, parental education, language spoken at home, and family income. Final Exam results also include a dummy variable identifying Institution B students who answered the common final questions in a follow-up data collection effort. Standard errors have been adjusted for clustering by section.*

TABLE A4. HYBRID EFFECTS ON LEARNING OUTCOMES, ROBUSTNESS CHECKS

	OLS/LPM Model		Exclude Non-Reg. Students		Excl. Inst. B	Std. by Inst.	Control Pre	Impute Post
	Complete	Pass	Complete	Pass	Final Exam	Final Exam	CAOS Post	CAOS Post
Hybrid	0.05	0.04	0.04+	0.03	0.02	0.19	0.01	0.02
	[0.03]	[0.04]	[0.02]	[0.03]	[0.03]	[0.14]	[0.01]	[0.02]
Observations	605	605	572	572	278	431	458	605

*Notes: ** p<0.01, * p<0.05, + p<0.1. All results control for institution dummies as well as student race/ethnicity, gender, age, full-time status, year in college, parental education, language spoken at home, and family income. Final Exam results also include a dummy variable identifying Institution B students who answered the common final questions in a follow-up data collection effort. Standard errors have been adjusted for clustering by section. "OLS/LPM Model" results use an OLS regression (linear probability model); "Exclude Non-Reg. Stud." excludes students who can be identified in the data as never registered for (enrolling in) the course: "Excl. Inst. B" excludes Institution B because two versions of the final exam were used; "Std. by Inst." standardizes the final exam percentage by institution to have a mean of zero and standard deviation of one; "Control Pre" controls for the pre-course CAOS score (which is set to zero for missing observations) and a dummy identifying students that do not have a pre-CAOS score; and "Impute Post" assigns students who did not take the post-CAOS their pre-CAOS score, or the mean pre-CAOS score at their institution if they did not take the pre-CAOS.*

TABLE A5. HYBRID EFFECTS ON LEARNING OUTCOMES, BY CAMPUS

	Complete	Pass	CAOS Post	Final Exam
Institution A	0.09	0.19+	-0.01	-0.12*
	[0.09]	[0.10]	[0.03]	[0.05]
	97	97	70	70
Institution B	0.05	0.04	-0.03+	0.08
	[0.05]	[0.05]	[0.02]	[0.05]
	229	229	189	153
Institution C	0.07	0.02	-0.00	-0.04
	[0.10]	[0.11]	[0.03]	[0.07]
	92	92	50	52
Institution D	0.15	-0.20	0.01	0.05
	[0.23]	[0.25]	[0.06]	[0.07]
	16	16	13	12
Institution E, Dept. 1	0.05	0.11	0.01	0.16
	[0.26]	[0.26]	[0.15]	[0.12]
	31	31	22	23
Institution E, Dept. 2	-0.07	-0.07	0.02	0.10*
	[0.05]	[0.05]	[0.05]	[0.05]
	50	50	44	45
Institution F	0.11	0.12	0.05	0.12**
	[0.09]	[0.08]	[0.04]	[0.04]
	90	90	70	76

*Notes: ** p<0.01, * p<0.05, + p<0.1. All results control for institution dummies as well as student race/ethnicity, gender, age, full-time status, year in college, parental education, language spoken at home, and family income. Final Exam results at Institution B also include a dummy variable identifying students who answered the common final questions in a follow-up data collection effort. Robust standard errors appear in brackets. Sample sizes appear in italics.*

TABLE A6. HYBRID EFFECTS ON LEARNING OUTCOMES, BY SUBGROUP

	Complete	Pass	CAOS Post	Final Exam
Black/Hispanic	-0.03	0.02	0.00	-0.00
	[0.05]	[0.05]	[0.01]	[0.04]
	188	188	143	131
White/Asian	0.10*	0.06	0.01	0.03
	[0.03]	[0.04]	[0.01]	[0.02]
	406	406	308	292
Male	0.05	0.04	-0.00	-0.00
	[0.05]	[0.06]	[0.02]	[0.03]
	257	257	194	173
Female	0.06+	0.05	0.01	0.04
	[0.03]	[0.04]	[0.01]	[0.03]
	348	348	264	258
Parents no college	0.03	0.01	-0.00	0.02
	[0.04]	[0.05]	[0.01]	[0.04]
	316	316	231	215
Parents college	0.07+	0.07	0.01	0.03
	[0.04]	[0.04]	[0.02]	[0.03]
	289	289	227	216
Fam. income <$50k	-0.02	-0.02	-0.01	-0.00
	[0.04]	[0.04]	[0.01]	[0.03]
	300	300	219	200
Fam. income ≥$50k	0.12**	0.11*	0.01	0.02
	[0.04]	[0.05]	[0.01]	[0.03]
	277	277	216	210
English only	0.06	0.07	0.01	0.03
	[0.04]	[0.05]	[0.01]	[0.02]
	384	384	289	283
English same/2nd	0.04	-0.00	-0.03*	-0.01
	[0.05]	[0.05]	[0.01]	[0.04]
	212	212	165	144
CAOS Pre low	0.01	0.03	0.01	-0.03
	[0.03]	[0.05]	[0.01]	[0.04]
	266	266	215	196
CAOS Pre high	0.01	-0.01	0.00	0.06+
	[0.02]	[0.03]	[0.01]	[0.03]
	265	265	234	222
Work less than 20 hours per week	0.04	0.04	0.00	0.02
	[0.04]	[0.04]	[0.01]	[0.03]
	431	431	329	311
Work more than 20 hours per week	0.09	0.06	-0.00	0.05
	[0.06]	[0.08]	[0.03]	[0.03]
	165	165	124	117

TABLE A6. HYBRID EFFECTS ON LEARNING OUTCOMES, BY SUBGROUP

	Complete	Pass	CAOS Post	Final Exam
College GPA less than 3.0	0.06	0.04	0.03*	0.00
	[0.04]	[0.06]	[0.01]	[0.04]
	314	314	221	219
College GPA 3.0 or higher	0.02	0.01	-0.00	0.05+
	[0.03]	[0.04]	[0.01]	[0.03]
	247	247	208	188

*Notes: ** p<0.01, * p<0.05, + p<0.1. All results control for institution dummies as well as student race/ethnicity, gender, age, full-time status, year in college, parental education, language spoken at home, and family income. Final Exam results also include a dummy variable identifying Institution B students who answered the common final questions in a follow-up data collection effort. Standard errors adjusted for clustering by section and appear in brackets. Sample sizes appear in italics.*

TABLE A7. DIFFERENCES IN STUDENT ASSESSMENT OF COURSE

	Overall	Interest	Learn	Difficulty	Hrs/Week
Hybrid	-0.25+	-0.04	-0.21+	0.22+	0.30
	[0.13]	[0.12]	[0.11]	[0.13]	[0.41]
Observations	435	440	438	440	437
Control Mean	2.3	1.7	2.2	2.3	4.0
Control SD	1.0	1.1	1.0	0.8	3.0

*Notes: ** p<0.01, * p<0.05, + p<0.1. All results control for institution dummies. Standard errors have been adjusted for clustering by section.*

Appendix B: Illustrative Cost Simulations

This appendix is intended to complement the discussion in the text of potential cost savings. It is to be read in conjunction with that discussion.

Important note: The data reported in this Appendix should be regarded as illustrative only. Our objective is to explore an approach to cost estimation and obtain very rough estimates of potential long-run savings, not to produce anything close to precise numbers.

The data used for the main part of our cost simulation analysis are instructor compensation data from three introductory statistics courses at two public universities in the Northeastern and Mid-Atlantic regions of the United States. Of the three statistics courses, one is offered as part of an undergraduate business program at one institution (Institution A), and the two other courses are offered in two different departments at a second institution (Institution B). The data are from the fall 2010 semester.

At Institution A, full professors are expected to teach seven three-credit courses each academic year with about 40 students enrolled in each. (Professors receive double teaching credit for a course with about 80 students.) Consequently, the compensation cost of teaching one section of a three-credit course is one-seventh of the annual wages and benefits for the faculty member. In fall 2010, average annual compensation of full professors who taught statistics at this institution was about $130,000, or about $19,000 per three-credit course. Other faculty members (generically called "part-time," and often adjuncts—see below for a further discussion of nomenclature) are compensated at an hourly rate that works out to be approximately $3,500 per three-credit course.

At Institution B, professors are expected to teach eight courses of about 25 to 35 students each academic year. In this setting, the compensation cost of a given section of introductory statistics is calculated as one-eighth of the annual wages and benefits of the faculty member. In fall 2010, the annual compensation of introductory statistics professors (averaged across the two different departments studied at this institution) was about $117,000 for full professors, $95,000 for associate professors, and $77,000 for assistant professors. These numbers correspond to per-course compensation of about $15,000, $12,000, and $10,000, respectively. Total compensation of "part-time" faculty was $3,600 in fall 2010.

The faculty compensation data are summarized in Table B1. Per-student compensation costs range from $425 to $450 for professors and from $101 to $147 for part-time faculty. In other words, compensation costs are roughly three to five times greater for tenure-track faculty than for part-time instructors. These large differences in compensation costs are a direct reflection of the fact that embedded "departmental research" costs are high for tenure-track faculty but low or non-existent for adjuncts and other part-time faculty. For example, at Institution A, professors and lecturers taught 29 percent of students in introductory statistics but received 64 percent of total compensation.

TABLE B1. INTRODUCTORY STATISTICS COMPENSATION COSTS, FALL 2010

	Total Sections	Total Students	Comp. per Student
Institution A			
Professors and Lecturers	4	234	$450
Part-time Faculty	9	575	$101
Total	13	809	$202
Institution B, Dept. 1			
Professors	8	238	$425
Part-time Faculty	4	98	$147
Total	12	336	$344
Institution B, Dept. 2			
Professors	4	107	$441
Part-time Faculty	8	204	$141
Total	12	311	$244

Notes: Compensation includes wages and benefits allocated to teaching. Part-time Faculty at Institution A include 1 staff member who taught part time. Institution A data exclude an honors section and an online section. Institution B, department 1 data exclude a partially online section and a section taught at an off-campus location.

There are many ways one could implement hybrid learning on a college campus. We focus on one model that seems plausible and includes a set of adjustable assumptions that make it quite flexible. We assume that students will learn mostly through machine-guided online systems such as those in the Carnegie Mellon introductory statistics course that was used in our empirical study of learning outcomes. Instead of attending class for 3 or 4 hours each week, as they do now in a traditional format, students instead attend a one-hour face-to-face session where they can ask questions and review concepts that they did not learn adequately through the online system.

In this hypothetical model, a full-time faculty member (usually a tenure-track professor) will be responsible for overseeing all sections of a large introductory course. The professor will be the faculty member of record for the class, and will be ultimately responsible for all academic aspects of the class (syllabus, exams, grading, etc.). Of course other instructors will assist with the actual implementation of tasks such as writing and grading exams—though in time we expect much grading to be done automatically (as in the grading models being developed now for some MOOCs, such as those offered by professors at places like MIT and Stanford). The professor will be assisted by a part-time instructor who will have administrative responsibilities for the entire course, such as scheduling and making sure that all students have ready access to the online part of the course.

Part-time instructors will be responsible for leading the weekly face-to-face meetings with students and (at present, pending further development of automated grading systems) for grading student assignments and exams. We should be clear that by "part-time instructors" we mean the group of instructors currently referred to using a variety of terms, including: adjuncts, part-time faculty, and contingent faculty. These individuals need not be employed part-time by the institutions—they could be full-time employees by virtue of teaching multiple sections of the same course (or different courses), but they are customarily paid per course taught. At institutions with graduate students, graduate teaching assistants could also fill this role.

In our basic model, we assume that the professor overseeing the course will receive teaching credit equal to two sections of a traditional, face-to-face course (of about 40 students at Institution A and 25 to 35 students at Institution B). We assume the part-time instructor with administrative responsibilities for the entire course will also receive compensation equivalent to two sections, although at the lower part-time rate. Finally, we assume that the part-time instructors leading the weekly face-to-face meetings will receive credit equivalent to one half-section of a traditional, face-to-face course. In other words, two hybrid sections are compensated in a similar fashion to one traditional section. The two hybrid sections will involve less total face-to-face time, but of course will involve grading more student assignments.

These starting assumptions (which were worked out in consultation with deans and others at our two case-study institutions) can easily be altered. We estimate the total compensation cost in our model as the total compensation of all instructors associated with the course, which varies with the amount of teaching credit that instructors receive, their compensation per teaching credit, and the size of the sections that meet weekly. Specifically, the total compensation cost is

Total comp = (*Prof credit*) × (*Prof comp per credit*) + (*Admin credit*) × (*Admin comp per credit*) + (*Number of sections*) × (*Adjunct credit*) × (*Adjunct comp per credit*),

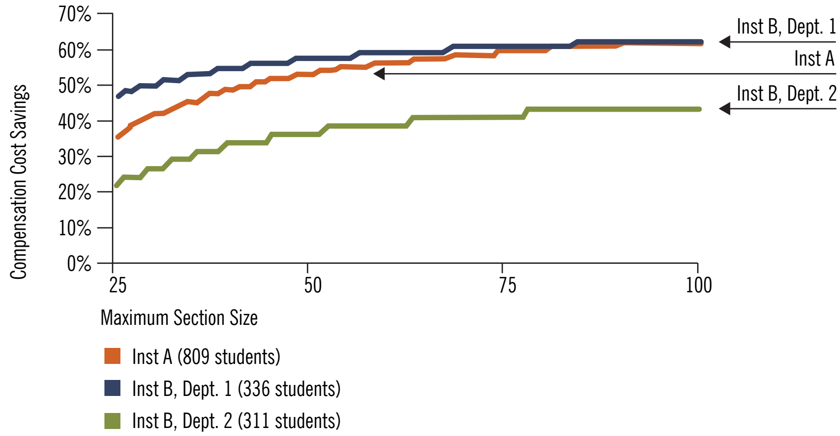
where the number of sections is defined by the ceiling function $\left\lceil \frac{\text{Enrollment}}{\text{Max. Section Size}} \right\rceil$.

(We have also constructed an Excel spreadsheet with macros that is intended to facilitate experimentation with alternative assumptions; see http://www.sr.ithaka.org.)

The compensation cost per student is calculated as the total compensation cost divided by the total enrollment of the course. For example, using the assumptions described above and a maximum section size of 50 students, the hybrid model at Institution A (with an enrollment of 809 students, the total enrollment in fall 2010) has compensations costs of $39,890 for the professor, $7,104 for the part-time administrator, and $30,192 for adjuncts responsible for leading 17 weekly face-to-face meetings. The total compensation cost of $77,186 is equal to $95 per student, which is $107 per student less than the current teaching model—a savings of 53 percent.

Our default assumptions yield predicted compensation cost savings of 36 percent in the statistics course in Department 2 and 57 percent in Department 1 of Institution B. Of course, using different assumptions in the model can change the estimated cost savings markedly. Figure B1 shows how estimated cost savings change when the maximum section size is changed from the default assumption of 50 to every possibility between 25 and 100. Cost savings are, of course, greater when sections are larger. However, there are still substantial cost savings even with section sizes in the 25 to 30 student range.

Figure B1. Compensation Cost Savings vs. Traditional Model



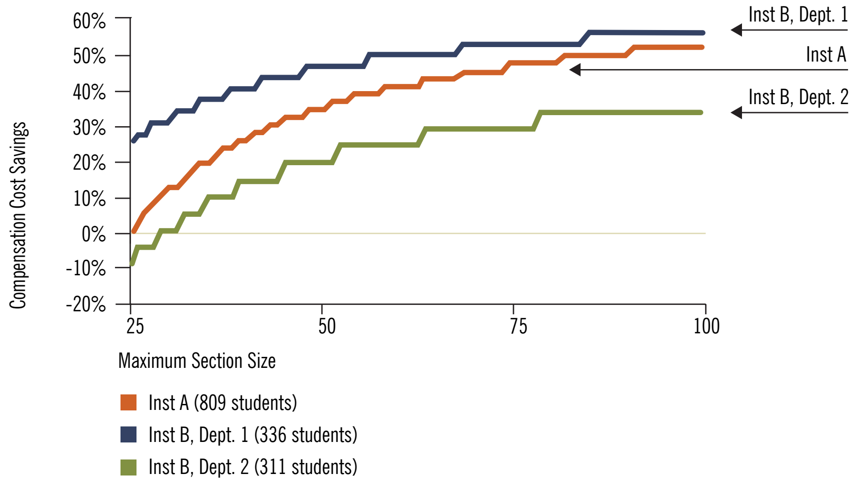
The reason that cost savings do not change that much with section size is because the biggest driver of compensation costs is the payment made to the professor in charge of coordinating the course. In the Institution A cost figures discussed above, the professor's compensation exceeds the combined compensation to the adjunct coordinator and the adjuncts responsible for 17 weekly in-person sessions.

One starting assumption that may deserve re-thinking is the assumption that part-time instructors will teach two sections to receive the same compensation they used to receive for teaching one section. The justification is that each hybrid section only entails one hour per week of class time instead of three or four. But the larger number of students means more assignments to grade and more students to keep track of (although the feedback system embedded in the online learning system may help in this regard). Independent of the question of how much teaching credit part-time instructors should receive is the question of what their (per credit) compensation should be. Some commentators have expressed concern that college students are increasingly being taught by a pool of poorly paid adjuncts who have to cobble together jobs at multiple institutions in order to eke out a living.

This larger question is outside the scope of this study. We can, however, examine how estimated cost savings change when the compensation of part-time instructors is doubled—which could be accomplished by doubling their teaching credit per section (from 0.5 to 1), doubling their compensation per credit, or some combination of an increase in teaching credit and an increase in compensation. Figure B2 shows that, in this simulation, significant cost savings are still realized in all three courses if section size is set at 40 to 50 or more, but only at one out of the three courses with a section size of 25 to 30.

The optimal hybrid teaching model will be different on each campus. Some campuses may prefer to put students in smaller sections and hire a larger number of instructors at a lower pay rate; others will prefer the opposite. Some campuses may be constrained by classrooms that are built for small classes, although this constraint may be less significant if not all students attend the weekly face-to-face sessions.

Figure B2. Compensation Cost Savings vs. Traditional Model, Double Adjunct Compensation

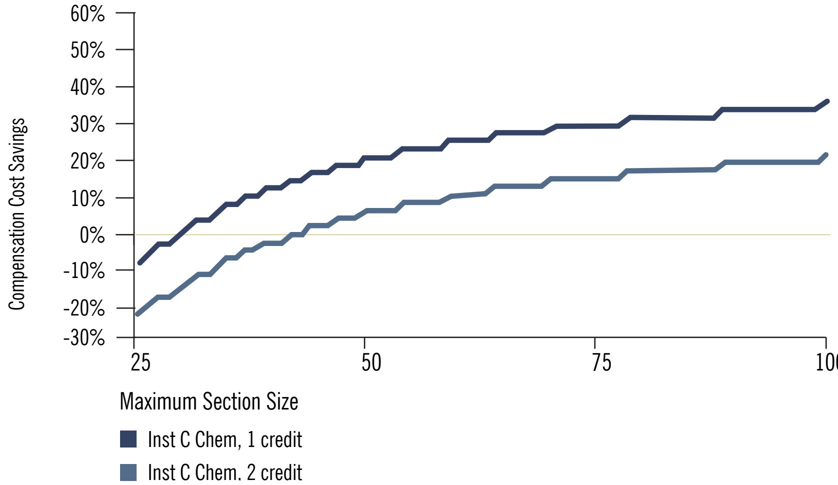


The two institutions that provided us with the compensation cost data referenced above use a traditional model of teaching in which students are taught in relatively small sections, some by professors and others by part-time instructors. The total compensation cost of instruction is driven largely by the share of instructors that are professors, since they are paid at a rate several times that of the adjuncts. Other institutions do not follow this model. Another common model is for a large introductory class to be taught in a large lecture that is supplemented by weekly meetings with teaching assistants.

This is the model used to deliver introductory chemistry instruction at a third institution with which we worked, Institution C. This institution provided us with estimated compensation of teachers instead of actual cost data. At Institution C we studied an introductory chemistry course that is taught in two lecture classes of 350 students each, for which the instructor receives compensation of $50,000 (or $25,000 per lecture "section"). Teaching assistants lead two sections of 72 students each and are paid $15,000 ($7,500 per section). There is also a full-time "discovery instructor" who provides extra assistance to students ($50,000 per year).

In this setting, the cost savings of a hybrid learning model relative to the traditional lecture-section model are lower because the full-time faculty costs are already spread over the entire class in the lecture-section model. At Institution C, if the hybrid course instructor is paid the same amount to serve as the academic coordinator for the course of 700 students as he or she would have been paid to teach a single lecture course of 350 students, cost savings are 19 percent (using all of the same assumptions discussed above). If instead the faculty member is paid for the two lectures that he or she used to teach, cost savings drop to 4 percent. Figure B3 shows the estimated cost savings for a range of section sizes. Significant cost savings are realized at the current traditional section size of 72 under both compensation scenarios, but there are no cost savings (and in some cases, cost increases) for smaller section sizes.

Figure B3. Compensation Cost Savings vs. Lecture-Section Model, Institution C Chemistry, by Professor Teaching Credit (Enrollment of 700)



Apart from these two basic teaching models, there are many other options. In the long run, institutions may not want to rely on the current pool of adjunct instructors available at current rates of pay. Instead, they might prefer to increase adjunct pay in order to attract individuals who are committed to teaching undergraduates and are glad to make a career doing so as long as they can make a decent living. A key is whether such individuals feel the need to be paid for some implied amount of "departmental research." Our simulations show that a hybrid learning model can decrease costs even if the instructors leading the face-to-face sessions are paid at a higher rate.

Once again, we wish to emphasize that all of the assumptions used in these illustrative simulations can be modified—and should be modified to suit local campus preferences. The simulations reported here are only "thought-starters."

Appendix C: Conducting a Randomized Study of Learning Outcomes on College Campuses
Lessons Learned from the Interactive Learning Online at Public Universities Project

Background

Between 2010 and 2012 Ithaka S+R coordinated a large-scale study of learning outcomes at nine public universities and community colleges (the Online Learning at Public Universities, or OLPU project). This report summarizes some of the lessons learned in conducting the study, with the goal of helping others who may be contemplating similar kinds of research.

In order to understand the research protocols described below, it is important to have a general understanding of the OLPU project. The overall purpose of our research was to assess the effectiveness of sophisticated, interactive modes of online learning. The study compared two methods of teaching introductory statistics. In the treatment group, students took the course in a "hybrid" mode, in which interactive, online course materials were supplemented with one hour of face-to-face instruction per week. Students in a control group took a traditional course, with 3-4 hours of lectures per week and standard textbooks. Students who agreed to participate in the study were randomly assigned to one of the two groups. Data were collected on learning outcomes, retention rates, and attitudes toward online instruction. Background data on students were also collected.

Many of the challenges associated with this study were related to finding effective ways to recruit as large a study population as possible, and then devising practical means to randomize the students into treatment and control groups. Various approaches were tried on different campuses. Below we outline the two approaches that we found to be the most effective. Both have their pros and cons. We do not believe there is a "one size fits all" approach; each campus has its own unique context, and each study will have different goals. Nonetheless, we believe there is much to be learned from our experiences, and that either of these approaches can be implemented successfully. We also provide some general considerations that have broad applicability to many different studies.

Coordinating the Study

As will be clear from the discussion that follows, this kind of study is complicated and involves the coordination of many different people and administrative units on campus. Based on our experience, the single most critical factor in making the project a success is identifying a strong project coordinator, or principal investigator (PI), on each campus. We recommend that the campus PI be paid a stipend commensurate with the amount of time required to manage the project. He or she should have the strong backing of senior administrators to help prod the campus bureaucracy when needed. The campus PI should also have enough administrative experience and seniority to interact with other administrative units, such as the Institutional Review Board, the registrar's office, the relevant academic department(s), the office of institutional research, and the dean and/or provost's office. In our observation—with some notable exceptions—asking a faculty member to play the role of campus PI is not the best choice because he or she may simply not be well enough connected with the other important units on campus. At the same time, asking too senior an administrator to take on this role is also not likely to be successful because of the amount of hands-on time needed. Finding the right person for the job is thus not a simple task.

We also strongly recommend convening all of the people involved in the study on each campus early in the planning process, including the instructors for both treatment and control sections. Getting everyone on board with an understanding of the overall purpose and design of the study can prevent a host of problems down the road. It is all too easy for someone to skip a step or take a shortcut, not realizing that doing so could compromise the success of the study. Ideally, all those who are investing significant time in the project above and beyond their normal duties, including instructors, administrators, and data analysts, should be compensated in some way, either financially or by being released from other duties.

Although, as described below, there are many advantages to having a pilot phase, one thing we did not anticipate was the amount of turnover that can take place in staffing a project that extends over the course of a whole year. In particular, some instructors who were involved in the pilot phase during the first semester of the study were replaced with new instructors in the following semester. This complicated the coordination of the study and meant that training had to be repeated, and that lessons learned in the pilot phase did not always carry over to the fall as smoothly as we had hoped.

Preparing for the Study

Given the complexity and number of players involved in this type of research, it is impossible to anticipate all potential problems. For this reason, we strongly recommend conducting a pilot study in advance; there is no substitute for a "dry run" to discover the hiccups and complications that inevitably arise—ranging from the mundane (not having enough copies of the consent forms on the first day of class) to the disastrous (mishandling student identifiers on the pre-course test, making the data unusable). The pilot study should be used as a learning process, with enough time planned to make adjustments based on what has been learned.[45]

In our case, the pilot phase involved implementing the full research protocol on every campus (except one) in the spring semester, prior to conducting the full study the following fall. Initially, we were tempted to run the pilot on just one or two campuses, assuming any lessons learned could be generalized across the other schools. However, we quickly found that the situations on each campus were so different that it was essential to run a full pilot on as many campuses as possible. In some cases, we worked with fewer sections or smaller groups of students during the pilot, but we deliberately did not skip any steps. In the best case, if everything went smoothly, we would have additional data to analyze. In the worst case, if the pilot was unsuccessful, we would have another chance to learn from our mistakes.

45 Full credit goes to James Kemple, Executive Director of the Research Alliance for New York City Public Schools, for initially recommending a pilot study.

Certain aspects of the study, highlighted below, required particularly long lead times.

- Before anything can begin, including advertising the study and recruiting participants, the campus Institutional Review Board (IRB) must give its approval (a process described in more detail below). Getting this approval can be a lengthy process because the IRB only meets at scheduled times, such as once a month, and may not meet at all during the summer or other breaks in the academic calendar. The IRB also frequently responds to the application by raising questions about the proposed protocol, and if changes are required, a revised plan may need to be re-submitted with a lengthy turn-around time. Given this unpredictability, we recommend beginning the process at least three or four months prior to the start of the study, keeping in mind the ebb and flow of the academic calendar.
- The research protocol may require recruitment of participants at the time students are registering for courses. It is important to remember that advance registration usually begins about four to six weeks before the end of the prior semester. Thus, fall registration may begin as early as late March. In addition, certain information may need to be changed in the campus's online course registration system, such as the addition of a new class section, a change to the maximum number of students permitted in a section, a change in a section time or location, or even a simple change to the description of a course. Typically, the registrar's office has difficulty making last-minute changes, and may not be accustomed to accommodating the needs of research studies. Thus, discussions with the registrar's office need to begin well before the start of registration.
- The OLPU project required multiple instructors on each campus who were willing to try a new online curriculum. Recruiting and then training the instructors cannot be done at the last minute. Teaching assignments may be set several months in advance, and instructors need extra time to become familiar with the new curriculum and software. We offered the training sessions remotely using webinars and conference calls and found that worked reasonably well and was much simpler to arrange than in-person sessions. Nonetheless, the logistics of arranging these sessions were often daunting and required adequate advance notice. We also found that it was helpful to have more than one training session—an initial one to allow the instructors to get started using the system, and then at least one follow-up session after they had tried it on their own.

For all of these reasons, it is desirable to begin planning a study like this as much as a year in advance.

Over the course of two semesters on nine campuses, a variety of approaches were used to recruit and randomize students into the treatment and control groups, all aimed at maximizing the number of participants. Below we describe the two approaches that seemed to be the most effective. No doubt there are other variations that would work as well, depending on the campus context and the goals of the study.

General Considerations

The biggest challenge in designing a protocol is to minimize the extent to which students who agree to participate and are randomized into one format (either the treatment or control group) subsequently decide they don't like their assignment and switch into the other format, or else drop the course altogether. Not only does this switching result in an unbalanced number of participants in the two groups, but, more seriously, it introduces a potential bias into the assignment process. To limit this problem, we recommend waiting to inform students about their assignment until as late as possible in the process, by which time they may be less likely to try to change their schedules. It also helps to have strong incentives to participate for all students in the study, with the clear message to students that they will lose the benefits of participation if they drop out or if they switch out of the format to which they were randomly assigned.

Careful scheduling of the study sections is critically important, unless the treatment group is purely online with no face-to-face instruction. Since students will not know which group they will end up in ahead of time, the treatment and control sections must be scheduled at the same time to ensure students can attend either section. It's also wise to select time slots that are known to be popular among students. Because the treatment and control sections meet at the same time (even if only one hour per week, as in the OLPU project), it is impossible to have the same instructor teach both the treatment and control sections.

Effective publicity of the study can make a big difference in the number of students who sign up. In addition to all the normal advertising channels (websites, posters, flyers, social media, etc.), it can be helpful to have information tables during registration, or schedule special information sessions with food as an incentive to attend.

Incentives

Offering meaningful incentives to persuade students to participate is essential to the success of the study. Taking an "experimental" course may be intimidating for many students, especially freshmen, who are often quite risk-averse to trying something new, as well as for students who have substantial doubts about how well they will do in the course. The stakes may seem especially high for a course that is a gateway to a desired course of study. Another factor may be bill-paying parents, who may be skeptical about the value of an online or hybrid course. Some students have had negative experiences with online courses in the past, or have heard horror stories from friends. For all of these reasons, without an effective set of incentives to make participation more attractive than non-participation, many students will take the path of least resistance and opt for the traditional mode of instruction.

In the OLPU project, campuses tried a variety of incentives: gift cards ranging in value from $10 to $50, keychains, MetroCards, free textbooks, and early registration the next semester. Our observation is that, while students seem to like gift cards or cash the most, small cash incentives (e.g. $10-20) did not seem to have much of an impact on students' decisions, especially at colleges where many students work full – or part-time and commute to campus (such incentives may work better at residential campuses where most students are full-time). A modest cash incentive may be adequate for a research study that asks students to commit a few hours or a day of their time, with no other long-term consequences, but for a decision that affects a whole course and (at least in some students' eyes) potentially their progress toward their major, the stakes are of a different order of magnitude. We found that $50 was the minimum amount needed to get students' attention. And incentives at that level probably only tip the scales for students who are on the fence about participating. We also learned that students prefer cash or general-purpose gift cards over specialized gift cards such as iTunes, Amazon, or the campus bookstore.

In designing incentives, it is important to make participation desirable to students regardless of whether they ultimately are assigned to the treatment or control group. In other words, the incentive should be perceived as meaningful to a participant regardless of the outcome of his or her random assignment. For example, in theory, students who were expecting to be in a traditional class are no worse off if they are randomly assigned to a control group. However, if they observe students in the treatment group receiving the equivalent of a free textbook, they may feel they are being treated unfairly. Or, if they were hoping to be assigned to a treatment group, the incentive will help lessen their disappointment and encourage their cooperation in completing the study requirements. Regardless, it is important the incentive not be distributed (or at least not fully distributed) until participants have fulfilled all the requirements of the study.

Option 1: Participants Sign Up on the First Day of Class

The simplest and most straightforward approach is to recruit students to participate on the first day of class. Ideally, three sections of the course should be scheduled at the same time—two of the sections will become the treatment and control groups, and the third will accommodate students who choose not to participate.[46] On the first day of class, students in all three sections should be instructed to go to the same meeting room.

After hearing an explanation of the study, students are invited to sign a consent form and immediately complete the baseline survey. (It can be highly effective to include an instant cash incentive in the form of a crisp ten dollar bill stapled to the consent form.) Any students who want more time to decide can return the signed consent form to a designated location within the next 24 hours. As soon as the deadline has passed, the campus PI randomly assigns students who have agreed to participate to either the treatment or control group. Participants are

46 If there are no caps on section size (or if the caps are very high), it is also possible to schedule only two concurrent sections, in which case the control group and non-participants can be combined in a single traditional section—provided that the non-participants and control group students can be distinguished for research purposes. However, we definitely recommend using three sections, if possible.

then sent an email informing them of the outcome and giving them instructions on when and where to meet for the next class. The campus PI also registers each student in the appropriate section in the registrar system.

The advantage of this approach is that it gives all students a thorough, in-person description of the study and an opportunity to ask questions. The presentation, which can include PowerPoint slides and a demonstration of the online materials, should be given by someone who is a dynamic presenter and has a good understanding of the study (the IRB will likely not allow the instructor to make the presentation). Most importantly, the "here and now" dynamic of the moment encourages students to make an immediate decision, and since their class schedules are already in place, they are less likely to change their minds or switch sections. If the study has been publicized in advance, some students may be already be inclined to participate.

The chief disadvantage of this approach is that the stakes are very high on that first day of class. If anything should go wrong—e.g. a snowstorm, technical difficulties, confusion about the room location—it is hard to recreate the opportunity. Or, if an outspoken student makes a negative comment about the study and sways the opinions of his peers, the participation rate can plummet. In addition, students who show up on the first day may still be "shopping" for classes, and other potential participants may not be present to hear the presentation.

The fact that there is uncertainty about the participation numbers until the very last minute can also create logistical problems for the campus and instructors. For example, suppose 80 students are present the first day and the participation rate is lower than expected, say only 30 percent. That means 24 participants will be randomly divided into treatment and control groups of 12 students each, leaving 56 non-participants. If that is too large for one section, some of the non-participants will have to be combined with the control group. Understandably, especially when classroom space is at a premium, not all administrators are comfortable with this degree of unpredictability or last-minute change. Some campuses do not have a large lecture hall available for the presentation on the first day, which then creates the complication of arranging for multiple presentations. Instructors also have to be willing to sacrifice a significant portion of one class period out of their course schedule.

Option 2: Participants Sign-Up During Online Course Registration

Our second recommended protocol avoids the "all eggs in one basket" problem but is somewhat more complicated to administer. In this approach, a single study section is set up in the campus's online course registration system. The section is reserved for study participants only, and is clearly labeled as such on the registration website. Ultimately, this section will be randomly divided into two sections, one for the treatment group and one for the control group. The initial section cap should thus be set high enough to accommodate the total number of desired participants.

Within 24 hours of registering (the sooner the better), students who sign up for the designated study section are sent an email by the campus PI with a link to an online consent form and baseline survey. Students who do not respond to the email within a set amount of time (such as a few days or a week) are warned by the project coordinator that they must agree to participate or they will be automatically dropped from the study section. If they do not respond by the deadline, they are "de-registered" from the study section, thereby freeing up spots for others to register. This process continues until the study section has filled with students who have agreed to participate. An important element of this approach is that students are not randomized right away. About one week before classes begin—or as close to the start of classes as possible—the campus PI conducts the randomization and informs students of the outcome.[47]

The main advantage of this approach is that it reduces the unpredictability of the recruitment process. The campus PI can monitor the number of students signing up throughout the registration process. If registrations are lower than expected, there is time to take other steps to publicize the study more aggressively. By the end of the registration period, the campus PI should have a pretty clear idea of the number of participants. By not informing students of their random assignments until as late as possible, the risk of students changing their minds or switching sections can be minimized. Unlike in Option 1, there is no need to arrange a third concurrent section for non-participants, nor is there a risk of ending up with unbalanced section sizes.

However, this approach has its own complications, both from the side of those who are running the study and from the student side. First, someone has to set up the online consent form and baseline survey, and make sure they are working properly throughout the registration process. Second, students may not carefully read the materials about the study and may not fully understand what they are signing up for, creating a potential for last-minute complaints or requests to switch to a different section. Third, adding an extra step to the registration process (having to complete the consent form and baseline survey) creates a hurdle that may limit the number of students who sign up, especially if they have many other section options from which to choose. Thus, this approach is unlikely to be successful without a strong incentive and extensive publicity. It also helps to schedule the study section during a popular time slot. Finally, this approach will not work without active management by the campus PI throughout the process—monitoring registrations, sending emails, answering questions, etc.

Data Collection

In the OLPU study, there were six separate sources of data collected: (1) the baseline survey administered to students at the time they agreed to participate; (2) a second student survey administered at the end of the semester; (3) a standardized test of statistical knowledge, administered once as a pre-course test and once as a post-course test; (4) final exam grades; (5) administrative data from the registrar's office or the institutional research office; and (6) an instructor questionnaire that gathered qualitative data describing the instructional practices used in, and the instructors' experience teaching, each study section.

47 A variation of this approach that some campuses tried is blocking students from registering online for the study section, and requiring them to sign-up in person at a designated place and time(s). This may work on a small campus or in a context where students are used to signing up in person, but there is a major risk that having to do so will be too big a hurdle for many students. However, this approach can work if students who want to take the class have to register in person.

Collecting these data was not as simple as we had hoped. The primary source of difficulty lay in correctly linking the participants' data from each source. This was complicated by the requirement that the data be stripped of all identifiers before being analyzed by the researchers. In theory, the campus PI and the registrar or institutional research office all made sure they were using the same anonymous study IDs, but it required careful coordination to make sure it was done correctly. Incorrect linking would make the results of the study meaningless. One simple way to guard against this potential problem is to include a few questions in the baseline survey that can be cross-checked with information in the administrative data as a way of confirming that the link has been done correctly. It is essential to work out a mechanism for saving the "linking file" in case researchers want to re-use the data in a future study.

It is critically important to carefully consider how the data instruments will be administered to students. The instructors are responsible for administering surveys, pre-course and post-course tests, and the final exam, so the researchers and campus PIs must provide clear instructions to ensure the data are collected consistently. When administering a pre-course or post-course test, all participants should all be given the same amount of time to complete the test. One reason for avoiding a study design in which participants and non-participants are mixed together is that it makes it more difficult to take time in class for the surveys and pre-course and post-course tests. It's important to schedule these instruments at a time when all participants are present. Ideally, the post-test should be given at the same time as the final exam. However, students (and instructors) may object to using time during the exam for that purpose, and asking participants to come early or stay late may be impractical.

One of the unexpected challenges we faced was making sure the treatment and control groups took a common final exam. Very few departments that we worked with required common final exams across all sections in the same class. As a result, in most cases, we had to negotiate with instructors to agree on a common set of final exam questions. It was surprisingly difficult to coordinate this. This was a critical element of the study, and there was only one opportunity to get it right—underscoring again the importance of running a pilot study to work out these critically important details before hand. A deeper issue was even finding consensus on what students were expected to learn in an introductory statistics course. In an ideal world, we would have preferred to develop a more carefully-targeted common assessment that we could have used across all campuses, but the practical problems of doing so were significant.

Institutional Review Boards

Because this kind of research involves "human subjects," it falls under the jurisdiction of the campus Institutional Review Board (IRB), which by law exists on all college campuses to protect the rights and welfare of participants in research projects. In general, the OLPU project did not encounter serious hurdles from IRBs on any of the campuses, although the process was often time-consuming and varied significantly from campus to campus in terms of the level of scrutiny and oversight. In about half the cases we qualified for "expedited review" (meaning that the IRB could approve the study without convening a meeting of the full IRB), and in the remaining cases we met the criteria for an "exempt" project[48] (meaning that the project was exempt from continuing review beyond the initial approval). However, the same approval process is required to determine whether a project is expedited or exempt. All researchers and principal investigators involved in the project were required to complete an approved IRB training course.[49] Given that we were working on nine different campuses, we engaged the help of an experienced outside legal consultant to help expedite the process.[50]

On more than one occasion an IRB requested clarification or minor changes to the proposed protocol. Their principal concerns generally fell into the following categories:

- *Protecting the anonymity of student data*: One issue was whether, in a class with a mix of participants and non-participants, the instructor would be allowed to know which students were participants. In practice, it was very difficult to prevent them from knowing. Because data were collected from multiple sources, it was important that at least one researcher on campus had access to student names and IDs to ensure the correct linking of data.
- *Ensuring students were not coerced into participating*: IRBs typically had two concerns. First, the instructor should not be the one who explained the study and invited students to participate. An unanticipated, and undesirable, consequence was that, on occasion, a person without a deep understanding of the study was expected to recruit students and answer their questions. The second concern was that the incentive not be set too high and thereby create an overwhelming incentive to participate. The incentives we used were deemed acceptable.
- *Providing information in clear, easy-to-understand language*: Some IRBs were appropriately concerned that the study be explained in simple terms without unnecessary "legalese" or fine print.
- *Allowing students to easily opt out of the study*: While an understandable and appropriate concern, in reality it was not practical for a student to switch out of a hybrid-format section into a traditional-format section after the first week or two of the course. The sequence of material and the manner of presentation were often too dissimilar. In addition, in order to maintain the integrity of the research design, it was not desirable to make it too easy for students to switch out of the section to which they had been randomly assigned simply because they did not like their assignment.
- *Providing background data and final course grades for non-participants*. In a few cases, IRBs inquired about our request for background data (demographic information, enrollment information, basic academic performance data, etc.)

48 There are several approved categories for exempt projects. The one that applied to the OLPU project was: "Research conducted in established or commonly accepted educational settings, involving normal educational practices, such as (a) research on regular and special education instructional strategies, or (b) research on the effectiveness or the comparison among instructional techniques, curricula, or classroom management methods."

49 The training is generally provided online through the Collaborative Institutional Training Initiative (CITI). Fortunately, the certificate provided by CITI was accepted by all of the campus IRBs.

50 We are extremely grateful to Jackie Ewenstein of Ewenstein & Young LLP for her assistance in this regard.

> and final course grades for students who did not agree to participate in the study. We were able to provide an adequate explanation of our intention to use these data for comparative purposes only, and to establish the legality of our request in accordance with the Family Educational Rights and Privacy Act. In addition, we agreed that no data whatsoever would be provided for students who not only decided not to participate in the study, but who also explicitly stated that they also did not want any of their data provided to researchers even for comparative purposes.

In all cases we were able to resolve the concerns raised by IRBs over these issues without undermining the integrity of the study. However, the discussions around these issues underscore the need for a clear and consistent understanding of which elements of the study are negotiable and which are not. This requires the involvement of an expert in research design throughout the approval process.

It is also important to point out that any significant changes to the research protocol—for example, at the end of the pilot phase, or in response to new developments—require approval from the IRB through a formal amendment process. Even changing the wording in a brochure or on a consent form can trigger the need for an amendment. This means that changes must be planned carefully in advance, and "bundled" together as much as possible, to minimize delays to the project.