INTERACTIVE LEARNING ONLINE AT PUBLIC UNIVERSITIES: EVIDENCE FROM A SIX-CAMPUS RANDOMIZED TRIAL

William G. Bowen, Ithaka S+R
Matthew M. Chingos, Brookings Institution and Ithaka S+R
Kelly A. Lack, Ithaka S+R
Thomas I. Nygren, Ithaka S+R

ABSTRACT

Online instruction is quickly gaining in importance in U.S. higher education, but little rigorous evidence exists as to its effect on student learning. We measure the effect on learning outcomes of a prototypical interactive learning online statistics course by randomly assigning students on six public university campuses to take the course in a hybrid format (with machine-guided instruction accompanied by one hour of face-to-face instruction each week) or a traditional format (as it is usually offered by their campus, typically with about three hours of face-to-face instruction each week). We find that learning outcomes are essentially the same—that students in the hybrid format pay no "price" for this mode of instruction in terms of pass rates, final exam scores, and performance on a standardized assessment of statistical literacy. We also conduct speculative cost simulations and find that adopting hybrid models of instruction in large introductory courses has the potential to significantly reduce instructor compensation costs in the long run.

INTRODUCTION

The American system of higher education is under increasing pressure to produce more graduates, and to do so with fewer resources. There is growing concern that the U.S. is losing its competitive edge in an increasingly knowledge-driven world, as many other countries make much more rapid progress than the U.S. in educating larger numbers of their citizens (Chingos, 2012). At the same time, higher education, especially in the public sector, is increasingly short of resources. Over the 10-year period from 2002 to 2012, state appropriations to their public universities decreased by 29 percent, on a per-student basis, from $8,352 to $5,906 in inflation-adjusted dollars. At the same time, enrollment increased by 28 percent, from 9.0 to 11.5 million full-time equivalent students (State Higher Education Executive Officers, 2013).

Falling support for higher education by state governments has been largely offset by increases in tuition revenue. But the days of higher tuition as an escape valve may be coming to an end, with growing concern about tuition levels and increasing resentment among students and their families that is having political reverberations. President Obama, in his 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 (Obama, 2012).

Today, a variety of higher education institutions must confront the challenge of how to manage costs in the face of tighter funding. In recent years, 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.[1] Moreover, a recent survey of provosts and chief academic officers found that very few of these administrators (and especially those at both public and private doctoral universities) gave their institutions high marks on effectiveness at controlling costs (Jaschik, 2012).

A fundamental source of the problem is the "cost disease," based on the labor-intensive nature of education with its attendant lack of opportunities for gains in productivity (Baumol and Bowen, 1966). 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

[1]According to the College Board (2011), tuition at public two-year universities in the 2011-2012 academic year increased, on average, by 8.7 percent relative to the previous academic year, a period during which tuition at public four-year institutions increased, on average, by 8.3 percent for in-state students and by 5.7 percent 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 percent for private not-for-profit four-year institutions and 3.2 percent for private for-profit institutions).

reductions in the cost of instruction.[2] Greater—and smarter—use of technology in teaching is widely seen as a promising way of controlling costs while also improving access. The exploding growth in online learning is often cited as evidence that, at last, technology may offer pathways to progress.[3] 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 (see, e.g., Christensen and Eyring, 2011).

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 settings. In important respects, the online learning marketplace reflects the diversity of American higher education itself.

The rapid growth in the adoption of online learning has been accompanied by an unfortunate lack of rigorous efforts to evaluate these new instructional models, in terms of their

[2] Baumol and Bowen (1966) argue 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). Consequently, unit labor costs will rise inexorably as these sectors have to compete for labor with other sectors in which productivity gains are easier to come by, and the relative costs of labor-intensive activities such as chamber music and teaching will therefore continue to rise. Bowen (2001) argues that, 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.

[3] A January 2013 report by the Babson Survey Research Group (Allen and Seaman, 2013) shows that between fall 2002 and fall 2011, 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 6.7 million, amounting to a compound annual rate of 17 percent (compared with a rate of three percent for course enrollments in general). More than three of every 10 students in higher education now take at least one course online.

effects on both quality and costs. There have been literally thousands of studies of "online learning," but the vast majority do not meet minimal standards of evidence (U.S. Department of Education, 2010) and only a handful involve semester-long courses in higher education (Jaggars and Bailey, 2010). Fewer still 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. And barely any studies use random assignment with sizeable student populations, leaving open the question of whether the results simply reflect student selection into online courses.

An important exception is Figlio, Rush, and Yin's (Forthcoming) randomized experiment in which they assigned students in an introductory microeconomics course at a selective research university to attend live lectures or watch online videos of the same lectures. They found no statistically significant differences in overall student achievement between the two formats, but did find evidence of negative online video effects among lower-achieving students, Hispanic students, and male students. There are several important differences between Figlio, Rush, and Yin's study and the present study that we return to below, but the most important distinction, in our view, is between the relatively primitive form of online instruction (videotaped lectures) evaluated in their study and the more sophisticated, interactive course examined in the present study (which we describe in more detail below).

Other studies comparing online and face-to-face formats involve still other variations of online or hybrid learning. The existing research, though subject to many caveats about quality and relevance, does not suggest that online or hybrid learning is more or less effective, on average, than traditional face-to-face learning (Lack, 2013). Not only do the types of online or hybrid learning involved in the literature vary considerably, but so do the kinds of outcomes measured, which range from homework assignment scores and project grades, to exam scores, final course grades, and completion and withdrawal rates. Many studies involve multiple measures of student performance, and within a single study, there are few instances in which one group outperforms the other group on all performance measures evaluated. The lack of consistency in findings may result from the wide variety of types of online learning studied and of research methodologies used, ranging from purely observational research to quasi-experimental studies to, in relatively few instances, randomized studies. Moreover, the variety in both research methodology and in forms of online learning, and the absence of a definitive pattern of online students consistently outperforming their face-to-face-format peers (or vice versa), render it difficult to reach any conclusions about what particular features of online courses are most or least conducive to enhancing student learning.

This study fills a significant gap in the literature about the relative effectiveness of different learning formats by providing the first evidence from randomized experiments of hybrid instruction conducted at a significant scale across multiple public university campuses. Given the pressing need for institutions to use limited resources as effectively as possible, the research reported here is concerned with educational costs as well, which have also received limited attention in prior research related to the effectiveness of online instruction.

We first describe the results of an experimental evaluation of 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. This section—which contains the results of the main part of this study—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.[4]

RESEARCH DESIGN

Our research is 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.

We worked with seven instances of a prototype ILO statistics course developed at Carnegie Mellon University (CMU).[5] The CMU statistics course 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.[6] 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;

[4] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://www3.interscience.wiley.com/cgi-bin/jhome/34787.
[5] The CMU statistics course can be accessed at http://oli.web.cmu.edu/openlearning/.
[6] Walsh (2011) describes the history of the development of this course, which was financed largely by the William and Flora Hewlett Foundation over a number of years.

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 in order to give the instructor and learning scientists a sense of how well students think they understand the concept.

However, although instructors can delete and re-order modules, CMU's platform 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-distant future.

Although 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 CMU statistics course was implemented in this hybrid format alongside traditional-format versions of the same course at six public university campuses (including two separate courses in two departments on one campus) that agreed to cooperate in a 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 listed in Table 1.

The exact research protocol varied by campus in accordance with local policies, practices, and preferences, and we describe these protocols in detail in Bowen et al. (2012). 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; 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 at the beginning of the semester; and 5) at the end of the semester, study participants were asked to take the CAOS test of statistical literacy again, as well as complete another questionnaire. The CAOS test, or Comprehensive Assessment of Outcomes in Statistics, is a 40-item, multiple-choice standardized assessment designed to measure students' statistical literacy and reasoning skills (delMas et al., 2007).

Administrative data on participating and non-participating students were gathered from the participating campus' institutional research offices. The baseline survey administered to students included questions on their 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, non-participants) enrolled in the statistics course in the fall 2011 semester.[7]

The treatment and control groups are described in Table 2. These data indicate that the randomization worked properly in that traditional- and hybrid-format students in fact have similar characteristics. Two differences are statistically significant at the 10 percent level, which is roughly what we would expect to find by random chance given that there are 22 characteristics examined. A regression of format assignment on all of the variables listed in Table 2 (and course dummy variables) fails to reject the null hypothesis of zero coefficients for all variables (except the course dummies) with p=0.16. A Hotelling test fails to reject the null of no difference in means with p=0.30.

Table 2 also shows that 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, 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.

Although the students participating in the study are a diverse group, they are a self-selected population because only students who agreed to be in the study were randomly assigned, and scheduling complications limited the population of participants. Overall, 605 of the 3,045 students enrolled in these statistics courses participated in the study. An even larger sample size would

[7] These data are described in detail in Bowen et al. (2012), and additional information including copies of the survey instruments is available at http://www.sr.ithaka.org/research-publications/interactive-learning-online-public-universities-evidence-randomized-trials.

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. 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 compared to other research on online learning.[8]

The data 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 89 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. Of course, the population of participants may be more likely to believe that they may benefit from a hybrid model of instruction. If that is the case, and if the hypothetical self-perception is accurate, then the hybrid course effects we estimate would be larger than we would obtain if we were able to randomly assign all students to a format without their consent. In a similar vein, the instructors who volunteered to teach the hybrid sections in this study may be particularly well-suited and motivated to teach in this format. Different results might be obtained if the hybrid sections were taught by instructors less well-suited to this mode of instruction.

[8] Of the 45 studies examined in the U.S. Department of Education (2010) meta-analysis, only five had sample sizes of over 400, and of the 50 independent effect sizes the authors abstracted, 32 came from studies with fewer than 100 study participants.

A notable limitation of these experiments is that although we were successful in randomizing students between treatment and control groups, we could not randomize instructors and thus could not control for differences in instructor quality. Table 4 reports student-weighted instructor and section characteristics by format for study participants. These data are drawn largely from instructor surveys, which were completed by instructors responsible for 90 percent of study participants. Table 4 shows that the hybrid sections were roughly similar in size to the traditional sections, but met for 1.5 hours less face-to-face time each week, on average. There were large differences in instructor characteristics, with hybrid instructors more likely to be employed full-time and to have taught online before but less likely to be tenure-track faculty. Hybrid instructors also had markedly less experience than traditional instructors, but still had 11 years of teaching experience, on average.

Some of the differences shown in Table 4 appear to advantage the hybrid sections, whereas others go in the opposite direction. Consequently, it is unclear a priori whether our results will overstate or understate the hybrid effect relative to an experiment that randomized both students and instructors. That depends not only on the balance of instructor characteristics in these experiments but also on the kinds of instructors who would be willing to be randomized to section format in the hypothetical experiment. We briefly return to this issue below and show that controlling for observable characteristics—an imperfect solution to this issue—does not qualitatively alter the results for three out of four of the learning outcomes we examine.

IMPACTS ON LEARNING OUTCOMES

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—four 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. The intent-to-treat (ITT) estimates that we report can be scaled up to treatment-on-the-treated (TOT) estimates by dividing by 0.04 (i.e. increasing the estimates by about four percent). We do not report TOT estimates because they are so similar to the ITT estimates, because most students took the course in the format to which they were randomly assigned.

Specifically, we estimate the following equation:

$$Y_{ic} = \beta_0 + \beta_1 Hybrid_{ic} + \delta_c + \epsilon_{ic},$$

where Y_{ic} is the outcome of student i in course c, β_0 is a constant, $Hybrid_{ic}$ is a dummy variable indicating whether the student was randomly assigned to the hybrid format (as opposed to the traditional format), δ_c is a vector of course-specific dummy variables, and ϵ_{ic} is the error term. We control for course dummies because 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 course dummies, as would be expected given that the probability of being assigned to the hybrid section was constant across courses (50 percent). The equation is estimated via ordinary least squares (OLS) for continuous outcomes and probit regression for binary outcomes (for the latter, we report marginal effects calculated at the mean of the independent variables). Standard errors are adjusted for clustering by course section in order to capture section-specific shocks to student outcomes (such as the quality of the instructor).[9]

[9] In some cases, students switched sections over the course of the semester. In these cases, we associated students with their section at the start of the semester if it was available in the administrative data. Students who were

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. Our main results are reported in Table 5.[10] The only statistically significant difference in learning outcomes between students in the traditional- and hybrid-format sections is the five-percentage-point higher course completion rate among the students assigned to the hybrid format. The difference in pass rates is slightly smaller, at four percentage points, and not statistically significant from zero. (A student can complete the course without passing it by remaining enrolled until the end of the semester but receiving a failing grade.) Hybrid-format students achieved similar scores on the CAOS test and slightly higher scores on the final exam.

We obtain similar results with and without including control variables, 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.

---

randomly assigned but never enrolled in the course are grouped as a "section" within each course for the purpose of computing clustered standard errors.

[10] Note that the pass rate in Table 5 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.

It is important to report that our estimated treatment impacts are fairly precisely estimated. We can be quite confident that treatment effects on course completion and pass rates were small or nil, and that the effects on CAOS and final exam scores were not large. For example, the results reported in the bottom panel of Table 5 indicate that we can rule out with 95 percent confidence the possibility that the hybrid format had a negative effect on pass rates of more than 2.4 percentage points. Likewise, we can rule out negative effects on CAOS and final exam scores larger than 1.6 and 2.8 percentage points, respectively. These 95 percent confidence interval lower bounds translate into 0.15 and 0.13 standard deviations (based on the distribution of the control group) for the CAOS and final exam, respectively. In other words, we can confidently rule out the possibility that assignment to the hybrid format had a large negative impact on student outcomes, but we cannot rule out small effects.

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.[11] We also conducted an analysis in which we grouped the 40 items on the CAOS test 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 each of the two groups of items.

Results that use final exam scores should also 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

[11] There was a larger increase in CAOS scores at the one campus where the test was part of student's final exam grade. In a study of 763 students at 20 institutions located in 14 states, the average increase was nine percentage points (delMas et al., 2007).

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 (analyses of final exam scores include a dummy variable identifying these students). At another institution, final exam data were not available for the students of two instructors (covering three out of six traditional-format sections).

A final potential concern with all of the outcomes besides the CAOS test is that they may have been affected by instructors' knowledge of the fact that they were part of a study, and of which students were study participants. For example, they may have used different grading standards that would affect pass rates. However, we think this is unlikely given that we obtain a nearly identical effect using completion rates, which are less likely to be affected by grading standards. The fact that we obtain a similar effect on CAOS and final exam scores increases our confidence that the effect on final exam scores is not biased by differential grading practices (which could not affect the multiple-choice CAOS test). In sum, despite the limitations of the each of the individual outcomes we examined, it is reassuring that the results are consistent across all of these outcomes.

Our results are also robust to a variety of alternative methodologies used to analyze the experimental data. These results are reported in Table A1.[12] First, we obtain nearly identical results when we use a linear probability modality (estimated via OLS) instead of a probit model to estimate treatment effects on binary outcomes. Second, we obtain similar results for completion and pass rates when we exclude students who agreed to participate in the study and were randomly assigned but never registered for the course. Third, we also obtain similar results when we exclude the institution where common final exam questions were administered in a follow-up data

[12] Tables A1, A2, and A3 appear in Appendix A. All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://www3.interscience.wiley.com/cgi-bin/jhome/34787.

collection, due to a lack of questions on the actual exam that were common across all six study sections at that campus.

Fourth, we obtain a larger and noisier estimate of the treatment effect on final exam scores when we standardize these scores separately by institution (instead of using percent correct as the outcome). Fifth, we obtain qualitatively similar results when we control for student scores on the CAOS pre-test (taken by 88 percent of participants), assigning a score of zero to students who did not take the test and identifying them using a dummy variable.

Sixth, 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 the estimated impact of assignment to the hybrid format on taking the CAOS and final exam is close to zero (not shown). 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.

Finally, we add controls for instructor characteristics, including full-time status, tenure-track status, years of teaching experience, and whether the instructor has taught online before, as well as section size (from the administrative records). Missing instructor characteristics (due to missing survey data) are assigned a value of zero, and these observations are identified using a dummy variable. Controlling for these characteristics is a crude approach to dealing with our inability to randomly assign instructors to section formats. The estimation of the relationship between these characteristics and student outcomes is likely to be very imprecise given the relatively small number of instructors in our data (we obtained survey data from 24 out of 26 instructors), and of course these models will not capture variation in instructor quality that is orthogonal to the measured characteristics. The final row of Table A1 shows that adding these controls leaves the general pattern of results unchanged. The estimated hybrid effects on completion and pass rates change sign, but are less precisely estimated. The effect on CAOS scores is approximately the same; only the effect on final exam scores changes sign and is statistically significant. Given the limitations of this approach to accounting for differences in instructor quality, we place more stock in the lack of an overall shift in results than in a significant change in one point estimate.

The lack of differences in mean outcomes between formats could mask differences in the distribution of outcomes. Figure 1 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).

Results broken down by individual institution (Table A2) 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.

We also calculated results separately for subgroups of students defined in terms of various characteristics, including race/ethnicity, gender, parents' education and income, 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 (Table A3).

The one possible 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 clear pattern for our other measure of socioeconomic status (parental education). Given the likelihood of finding spurious effects when a large number of coefficients are estimated (as in Table A3), the most likely conclusion is that there were no groups of students who benefited from or were harmed by the hybrid format consistently across multiple learning outcomes.

This conclusion differs noticeably from Figlio, Rush, and Yin's (Forthcoming) finding of negative effects of watching video-taped lectures (relative to live lectures) among Hispanic, male, and lower-achieving students. As discussed earlier, the most important difference between that study and the present one is the very different type of technology evaluated: video-taped lectures as compared to a sophisticated, interactive online course. There are at least three other differences that may also account for the difference in findings. First, the contexts were quite different: an economics course at a highly selective university as opposed to statistics courses at moderately selective universities. Second, Figlio, Rush, and Yin compared an online-only format to a live-only format (neither had discussion sections), whereas the present study compares a hybrid format to a live-only format (although courses in both studies had web sites for the distribution of course materials). Third, the different formats in Figlio, Rush, and Yin's study were taught by the same instructor, whereas different instructors taught the hybrid and traditional formats in the present study. But taken together, our results and those in Figlio, Rush, and Yin indicate that a more expensive hybrid course may yield better outcomes than simply presenting traditional large lecture courses in an electronic medium, a strategy that universities may pursue as a cost-cutting device.

In addition to examining learning outcomes, we also asked students how much they liked the course that they took (Table 6). We found that students gave the hybrid format a modestly lower overall rating than the one given by students taking the course in the traditional format (the rating was about 11 percent lower). By similar margins, hybrid students reported feeling that they learned less and that they found the course more difficult.[13] 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.

Finally, we 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 difference, which is not statistically significant, implies that, in a course where the traditional section meet for 1.5 hours more per week than the hybrid sections (see Table 4), the average hybrid-format student would spend 1.2 less hours each week in *total time* devoted to the course, a difference of about 18 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 (Lovett, Meyer, and Thille, 2008).

In sum, our results indicate that hybrid-format students took about one-fifth less time to achieve essentially the same learning outcomes as traditional-format students. The three main limitations of this analysis are: 1) the fact that we were not able to randomly assign instructors to

[13] 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 or been more 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.

section formats—which would have been difficult, if not impossible, to do; 2) the limits to external validity that result from the need to recruit students willing to be randomized; and 3) the limitations of the CMU prototype of an ILO course. Despite these limitations, these results represent the first rigorous assessment of the relative efficacy of technology-enhanced learning (ILO-style hybrid instruction) compared to the traditional mode of instruction in large introductory courses on multiple public university campuses.

At a minimum, this study supports a "no-harm-done" conclusion regarding one current prototype of an ILO system. But there is, without doubt, much more research that can and should be carried out. Future experimental studies should examine courses in subjects other than statistics. ILO courses may have more potential in subject areas where there is usually a "right answer," such as math and science, but there is little evidence on this question. Future research should also examine the intersection between the instructor and the delivery method. Having the same instructor teach in both formats would allow for an evaluation that holds constant the quality of the instructor, and if repeated over time would produce evidence as to whether some instructors are more effective when teaching in a particular format.

COSTS AND POTENTIAL SAVINGS

The experimental data on learning outcomes results described above show that a relatively sophisticated prototype hybrid learning system did not lead to a significant increase in outputs (student learning), but could potentially increase productivity by using fewer inputs. Costs are difficult to measure at the course level, which is a likely reason why so few prior studies have paid much attention to costs.[14] The key problem is that contemporaneous comparisons can be near-

[14] Carol Twigg's work with the National Center for Academic Transformation (NCAT) project is a notable exception (see the NCAT website at www.thencat.org).

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. This 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 amortized over time. There are also transition costs entailed in moving from the traditional, mostly face-to-face model 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 these problems, we made a very rough attempt to simulate cost savings 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 a 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 simple cost models—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. These simulations, at the minimum, provide at least a rough sense of the potential impact on costs of introducing hybrid learning and, more specifically, indicate the extent of opportunities for institutions to share cost savings with faculty and students on a continuing basis.

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 conducted 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).

These simulations, which are described in detail in Appendix B, 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. These simulations illustrate that hybrid learning offers opportunities for significant savings in compensation costs, but that the degree of cost reduction depends (as one might expect) on the exact model of hybrid learning used—especially the rate at which instructors are compensated and section size (as detailed in Appendix B). 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 question-and-answer settings. Of course, tenured professors cannot be laid off in order to realize these savings; 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. Hybrid models make it possible to teach more students without increasing the demands made on tenured faculty. Recruitment costs may thereby be reduced along with compensation costs per student, avoiding debates over maintaining commitments to existing faculty. 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 classroom 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 classroom 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. Table 4 indicates that the hybrid course is scheduled to meet for 1.2 hours each week, as compared to 2.7 hours for the typical traditional course. Consequently, the hybrid course requires 56 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.

This highly speculative cost simulation effort cannot provide precise predictions of cost savings from a shift to hybrid learning, but strongly suggests that this pedagogical model has the potential to improve educational productivity by achieving equivalent learning outcomes at reduced costs.

CONCLUSIONS

In the case of a topic as active as online learning, where millions of dollars are being invested by a wide variety of entities, we might expect 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. Vigorous efforts should be made to continue to explore and assess uses of both the relatively simple systems that are proliferating, 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, and thus it is not unreasonable to speculate that learning outcomes will also improve.

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. ILO systems can also improve educational productivity by producing equivalent learning outcomes at a reduced cost. Furthermore, by (potentially) saving significant amounts of resources, such systems could lead to more, not less, opportunity for students to benefit from exposure to modes of instruction such as directed study—if scarce faculty time could be beneficially redeployed.

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. The evidence presented in this paper suggests 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.

REFERENCES

Allen, I. E., & Seaman, J. (2013). Changing course: Ten years of tracking online education in the United States. Babson Park, MA: Babson Survey Research Group.

Baumol, W. J., & Bowen, W. G. (1966). Performing arts: The economic dilemma. New York: Twentieth Century Fund.

Bowen, W. G. (2001). At a slight angle to the universe: The university in a digitized, commercialized age. Princeton, NJ: Princeton University Press.

Bowen, W. G., Chingos, M. M., Lack, K. A., & Nygren, T. I. (2012). Interactive learning online at public universities: Evidence from randomized trials. New York: Ithaka S+R.

Chingos, M. M. (2012). Graduation rates at America's universities: What we know and what we need to know," in Andrew P. Kelly and Mark Schneider, eds., Getting to graduation: The completion agenda in higher education. Baltimore: Johns Hopkins University Press, pp. 48–71.

Christensen, C. M., & Eyring, H. J. (2011). The innovative university: Changing the DNA of higher education from the inside out. San Francisco: Jossey-Bass.

College Board. (2011). Trends in College Pricing 2011. New York: College Board.

delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. Statistics Education Research Journal, 6, 28-58.

Figlio, D. N., Rush, M. & Yin, L. (Forthcoming). Is it live or is it internet? Experiment estimates of the effects of online instruction on student learning. Journal of Labor Economics.

Jaggars, S. S., & Bailey, T. (2010). Effectiveness of fully online courses for college students: Response to a department of education meta-analysis. New York: Community College Research Center, Teacher's College, Columbia University.

Jaschik, S. (2012). Mixed grades: A survey of provosts. Inside Higher Education, January 25.

Lack, K. A. (2013). Current status of research on online learning in postsecondary education. New York: Ithaka S+R.

Lovett, M., Meyer, O., & Thille, C. (2008). The Open Learning Initiative: Measuring the effectiveness of the OLI statistics course in accelerating student learning. Journal of Interactive Media in Education.

Obama, Barack. (2012). Remarks by the president in state of the union address. Washington, DC: White House, January 24, 2012.

State Higher Education Executive Officers. (2013). State higher education finance FY 2012. Boulder, CO: Author.

U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. (2010). Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies. Washington, DC: Author.

Walsh, T. (2011). Unlocking the gates: How and why leading universities are opening up access to their courses. Princeton, NJ: Princeton University Press.

APPENDIX A

[Tables appended to end of PDF.]

APPENDIX B: ILLUSTRATIVE COST SIMULATIONS

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 in an 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.

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 about three 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.). 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 massively open online courses, 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 for the hybrid courses will receive credit equivalent to one half-section of a traditional, face-to-face course. In other words, part-time instructors are compensated the same amount for two hybrid sections as for one traditional section. The two hybrid sections will involve less total face-to-face time, but 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

$$\textit{Total comp} = (\textit{Prof credit}) \times (\textit{Prof comp per credit}) + (\textit{Admin credit}) \times (\textit{Admin comp per credit}) + (\textit{Number of sections}) \times (\textit{Adjunct credit}) \times (\textit{Adjunct comp per credit}),$$

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

(We have also constructed an Excel spreadsheet with embedded formulas 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 compensation costs of \$39,890 for the professor, \$7,104 for the part-time administrator, and \$30,192 for adjuncts responsible for leading 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 1 and 57 percent in Department 2 of Institution B. As one might expect, 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.

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 of the adjunct coordinator and the adjuncts responsible for weekly in-person sessions of 17 course sections.

One starting assumption that may deserve re-thinking is the assumption that part-time instructors will teach two hybrid sections to receive the same compensation they used to receive for teaching one traditional 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 may 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.

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 who 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 instructors 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 than the cost savings when the all-section model is used, because in the lecture-section model the full-time faculty costs are already spread over the entire class. 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 lecture courses that he or she used to teach, cost savings drop to 4 percent (still assuming that the two traditional lecture courses each had 350 students). 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, there are cost increases) for smaller section sizes.

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.

Figure 1. Distributions of CAOS Post-Test Scores

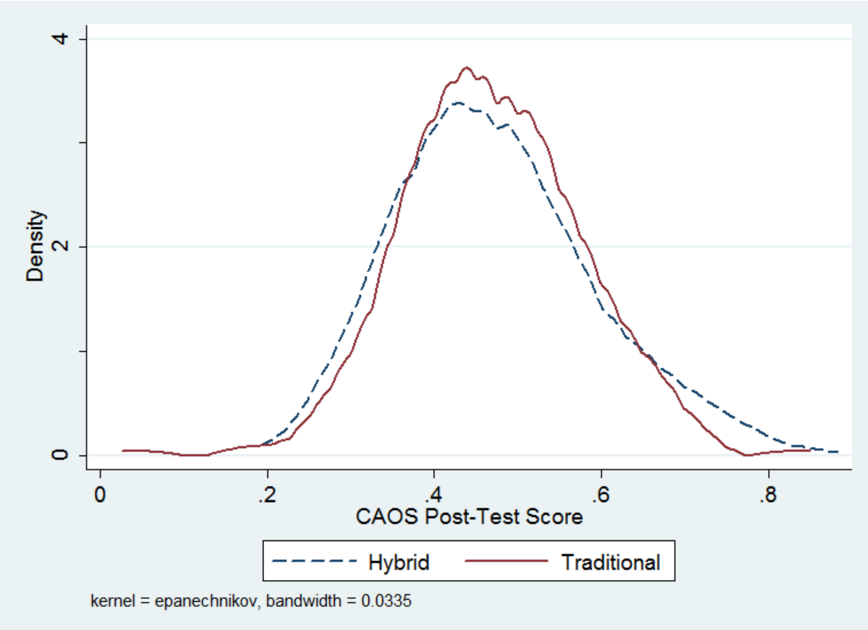

Figure B1. Compensation Cost Savings vs. Traditional Model

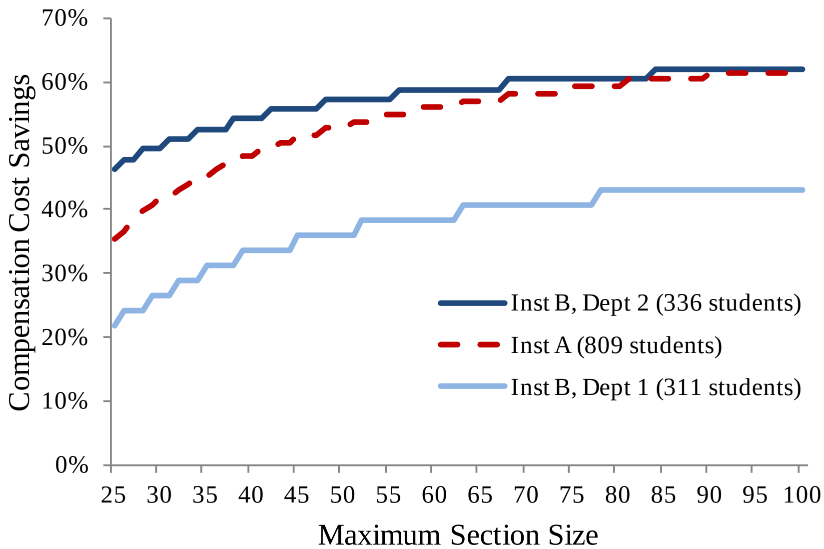

Figure B2. Compensation Cost Savings vs. Traditional Model, Double Adjunct Compensation

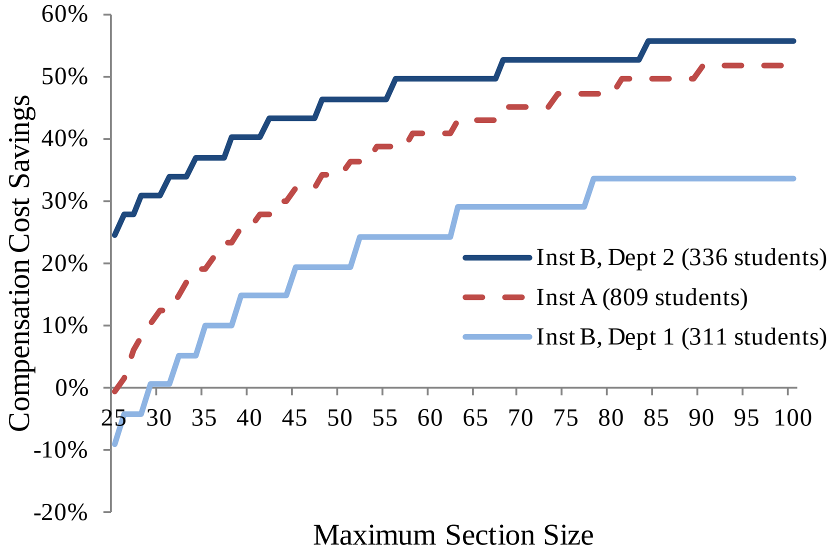

Figure B3. Compensation Cost Savings vs. Lecture-Section Model, Institution C Chemistry, by Professor Teaching Credit (Enrollment of 700)

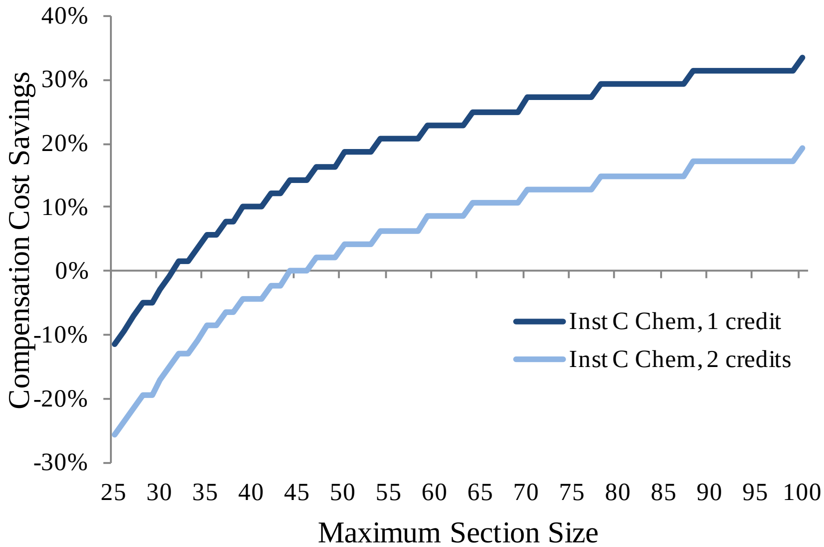

Table 1. Participating Courses/Institutions, Fall 2011

	Course Enrollment	Study Participants
Institution A	850	97
Institution B	86	16
Institution C	876	229
Institution D	235	92
Institution E, Dept 1	337	31
Institution E, Dept 2	473	50
Institution F	188	90
Total	3,045	605

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

Table 2. Randomization of Study Participants

	Traditional	Hybrid	Adj Diff	p-value
Male	0.46	0.39	-0.07	0.08
Race/Ethnicity				
Asian	0.24	0.23	-0.01	0.82
Black	0.14	0.14	0.00	0.99
Hispanic	0.20	0.14	-0.05	0.07
White	0.41	0.46	0.04	0.24
Other/Missing	0.01	0.03	0.02	0.16
Age				
Average	21.9	22.0	0.0	0.92
Less than 24	0.82	0.84	0.02	0.57
Between 24 and 30	0.14	0.10	-0.04	0.16
30 or greater	0.04	0.05	0.01	0.53
College GPA				
Missing	0.10	0.08	-0.03	0.27
Less than 2.0	0.09	0.08	-0.01	0.53
Between 2.0 and 3.0	0.37	0.40	0.03	0.37
3.0 or greater	0.44	0.44	0.00	0.91
Enrolled full-time	0.89	0.88	-0.01	0.64
Year in college				
Freshman	0.11	0.09	-0.02	0.37
Sophomore	0.42	0.47	0.05	0.25
Junior	0.35	0.31	-0.03	0.38
Other/Missing	0.12	0.13	0.00	0.89
Family income < $50,000	0.49	0.50	0.02	0.63
Parent college graduate	0.49	0.47	-0.02	0.64
English only language	0.65	0.62	-0.04	0.25
N	292	313		

Notes: Adjusted differences control for course dummies; the p-value of the adjusted difference is also listed (based on robust standard errors). A regression of format assignment on all variables listed here fails to reject null of zero coefficients for all variables with p=.16.

Table 3. Student Characteristics by Study Participation

	Participant	Non-Part.	Adj Diff	p-value
Male	0.42	0.44	-0.01	0.81
Race/Ethnicity				
Asian	0.23	0.17	0.01	0.58
Black	0.14	0.13	0.00	0.85
Hispanic	0.17	0.10	0.03	0.04
White	0.44	0.47	0.06	0.01
Other/Missing	0.02	0.13	-0.10	0.00
Age				
Average	21.9	21.6	-0.3	0.14
Less than 24	0.83	0.81	0.05	0.01
Between 24 and 30	0.12	0.10	-0.01	0.37
30 or greater	0.05	0.04	-0.01	0.38
College GPA				
Missing	0.09	0.15	-0.06	0.00
Less than 2.0	0.08	0.10	-0.01	0.50
Between 2.0 and 3.0	0.39	0.37	0.01	0.64
3.0 or greater	0.44	0.38	0.06	0.01
Enrolled full-time	0.89	0.84	0.04	0.01
Year in college				
Freshman	0.10	0.18	-0.05	0.00
Sophomore	0.45	0.40	0.04	0.11
Junior	0.33	0.27	0.02	0.25
Other/Missing	0.13	0.15	-0.01	0.50
Passed course	0.78	0.81	-0.04	0.03
Completed course	0.84	0.87	-0.04	0.01
Course grade	2.37	2.36	-0.03	0.61
N	605	2,440		

Notes: Adjusted differences control for course dummies; the p-value of the adjusted difference is also listed (based on robust standard errors). Students that did not complete course are assigned a course grade of zero.

Table 4. Instructor and Section Characteristics, by Format

	Traditional	Hybrid	Adj Diff	p-value
Face-to-face minutes scheduled	163	70	-92	0.00
Section size (administrative data)	67.0	74.7	5.4	0.01
Section size (instructor survey)	76.9	72.8	-2.5	0.18
Full-time instructor	0.34	0.50	0.21	0.00
Tenure-track instructor	0.23	0.16	-0.04	0.14
Years of teaching experience	20.2	11.1	-8.2	0.00
Number times taught intro stats	49.6	20.7	-26.5	0.00
Number times taught this course	40.2	16.5	-22.1	0.00
Taught online before	0.32	0.75	0.46	0.00

Notes: Summary statistics are weighted by student enrollment (study participants only). Number of student observations is 605 for section size from administrative data, 561 for scheduled face-to-face minutes, and 543 for all other characteristics (from instructor survey). Adjusted differences control for course dummies; the p-value of the adjusted difference is also listed (based on robust standard errors).

Table 5. Hybrid Effects on Learning Outcomes

	Without Controls			
	Complete	Pass	CAOS Post	Final Exam
Hybrid	0.05*	0.04	0.00	0.03
	[0.02]	[0.03]	[0.01]	[0.03]
Observations	605	605	458	431

	With Controls			
	Complete	Pass	CAOS Post	Final Exam
Hybrid	0.05+	0.04	0.01	0.02
	[0.03]	[0.03]	[0.01]	[0.02]
Observations	605	605	458	431
Control mean	0.82	0.76	0.47	0.55
Control std dev	-	-	0.11	0.22

Notes: * $p<0.05$; + $p<0.1$. Standard errors adjusted for clustering by section appear in brackets. Results for Complete and Pass rates are marginal effects from probit regressions (calculated at the mean of the independent variables). All results control for course 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 A students who answered the common final questions in a follow-up data collection effort.

Table 6. 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 std dev	1.0	1.1	1.0	0.8	3.0

Notes: + p<0.1. Standard errors adjusted for clustering by section appear in brackets. All results control for course dummies.

Table A1. Hybrid Effects on Learning Outcomes, Robustness Checks

	Complete	Pass	CAOS Post	Final Exam
Main estimates with controls (from Table 5)	0.05+	0.04	0.01	0.02
	[0.03]	[0.03]	[0.01]	[0.02]
	605	*605*	*458*	*431*
OLS/LPM model	0.05*	0.04		
	[0.03]	[0.03]		
	605	*605*		
Exclude non-registered students	0.03	0.02		
	[0.02]	[0.03]		
	563	*563*		
Exclude Institution A	0.06	0.04	0.02+	0.02
	[0.03]	[0.05]	[0.01]	[0.03]
	376	*376*	*269*	*278*
Standardize final exam scores by institution				0.19
				[0.14]
				431
Control for CAOS pre-test scores (set to zero if missing, incl dummy)	0.02	0.02	0.01	0.02
	[0.02]	[0.03]	[0.01]	[0.02]
	605	*605*	*458*	*431*
Impute CAOS post-test scores using pre-test scores			0.02	
			[0.01]	
			605	
Control for instructor characteristics and section size	-0.03	-0.03	-0.00	-0.05*
	[0.04]	[0.06]	[0.01]	[0.02]
	605	*605*	*458*	*431*

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

Table A2. Hybrid Effects on Learning Outcomes, by Campus

	Complete	Pass	CAOS Post	Final Exam
Institution A	0.10	0.22*	0.01	-0.11*
	[0.10]	[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,	0.05	0.11	0.01	0.16
Department 1	[0.26]	[0.26]	[0.15]	[0.12]
	31	*31*	*22*	*23*
Institution E,	-0.07	-0.07	0.02	0.10*
Department 2	[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: ** $p<0.01$; * $p<0.05$; + $p<0.1$. Robust standard errors appear in brackets. Sample sizes appear in italics. All results control for course 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 A also include a dummy variable identifying students who answered the common final questions in a follow-up data collection effort.

Table A3. Hybrid Effects on Learning Outcomes, by Subgroup

	Complete	Pass	CAOS Post	Final Exam
Black/Hispanic	-0.02	0.02	-0.00	-0.00
	[0.04]	[0.04]	[0.01]	[0.04]
	188	*188*	*143*	*131*
White/Asian	0.10**	0.07	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	0.07*	0.06	0.01	0.04
	[0.03]	[0.03]	[0.01]	[0.03]
	348	*348*	*264*	*258*
Neither parent has a	0.03	0.02	-0.00	0.02
bachelor's degree	[0.04]	[0.04]	[0.01]	[0.04]
	316	*316*	*231*	*215*
At least one parent has	0.07+	0.07	0.01	0.03
a bachelor's degere	[0.04]	[0.04]	[0.02]	[0.03]
	289	*289*	*227*	*216*
Parents' income	-0.02	-0.02	-0.01	-0.00
less than $50,000	[0.03]	[0.03]	[0.01]	[0.03]
	300	*300*	*219*	*200*
Parents' income	0.13**	0.12*	0.01	0.02
$50,000 or more	[0.04]	[0.05]	[0.01]	[0.03]
	277	*277*	*216*	*210*
English only language	0.06+	0.08	0.01	0.03
spoken	[0.03]	[0.05]	[0.02]	[0.02]
	384	*384*	*289*	*283*
English spoken in addition	0.05	-0.00	-0.03*	-0.01
to other language	[0.05]	[0.05]	[0.01]	[0.04]
	212	*212*	*165*	*144*
Pre-CAOS in bottom half	0.01	0.03	0.01	-0.03
	[0.03]	[0.05]	[0.01]	[0.04]
	266	*266*	*215*	*196*
Pre-CAOS in top half	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	0.04	0.04	0.00	0.02
per week	[0.03]	[0.04]	[0.01]	[0.02]
	431	*431*	*329*	*311*
Work more than 20 hours	0.09	0.07	-0.00	0.05
per week	[0.06]	[0.08]	[0.03]	[0.03]
	165	*165*	*124*	*117*
	0.03	0.01	0.03*	0.04
College GPA less than 3.0	[0.04]	[0.06]	[0.01]	[0.03]
	284	*284*	*194*	*192*
	0.03	0.02	-0.00	0.02
College GPA 3.0 or higher	[0.04]	[0.04]	[0.01]	[0.03]
	266	*266*	*226*	*206*

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

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	4	107	$441
Part-Time Faculty	8	204	$141
Total	12	311	$244
Institution B, Dept. 2			
Professors	8	238	$425
Part-Time Faculty	4	98	$147
Total	12	336	$344

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 psychology data exclude a partially online section and a section taught at an off-campus location.