Journal of Labor Economics

# Is it Live or is it Internet? Experimental Estimates of the Effects of Online Instruction on Student Learning

--Manuscript Draft--

| | |
|---|---|
| **Manuscript Number:** | 9226R4 |
| **Full Title:** | Is it Live or is it Internet? Experimental Estimates of the Effects of Online Instruction on Student Learning |
| **Article Type:** | Article |
| **Corresponding Author:** | David Figlio<br>Northwestern University<br>Evanston, IL UNITED STATES |
| **Corresponding Author Secondary Information:** | |
| **Corresponding Author's Institution:** | Northwestern University |
| **Corresponding Author's Secondary Institution:** | |
| **First Author:** | David Figlio |
| **First Author Secondary Information:** | |
| **Order of Authors:** | David Figlio |
| | Mark Rush |
| | Lu Yin |
| **Order of Authors Secondary Information:** | |
| **Manuscript Region of Origin:** | USA |
| **Abstract:** | This paper presents the first experimental evidence on the effects of live versus internet media of instruction. Students in a large introductory microeconomics course at a major research university were randomly assigned to live lectures versus watching these same lectures in an internet setting, where all other factors (e.g., instruction, supplemental materials) were the same. We find modest evidence that live-only instruction dominates internet instruction. These results are particularly strong for Hispanic students, male students, and lower-achieving students. We also provide suggestions for future experimentation in other settings. |

# Is it Live or is it Internet?

# Experimental Estimates of the Effects of Online Instruction on Student Learning

by

David Figlio
Northwestern University, NBER and CESifo

Mark Rush
University of Florida

Lu Yin
University of Florida

Revised: March 2011

The authors are grateful to the anonymous university and microeconomics instructor for providing registrar data and for permitting the experiment to take place and facilitating the experiment. We benefited greatly from comments from seminar participants at Northwestern University, the University of Florida, and the University of Michigan, and audience members at the annual meeting of the American Economic Association. We are grateful to Susan Dynarski and the anonymous referees for very helpful comments. All remaining errors are our own.

Throughout the United States, public four-year colleges and universities are facing fiscal constraints not seen in four decades. State and local appropriations for higher education, measured as a share of personal income, have fallen virtually monotonically since the late 1970s, and today are at a level not seen since the mid-1960s (Mortenson, 2005). Kane and Orszag (2003) document the precipitous decline in per student spending and stature of public four-year colleges and universities during the 1980s and 1990s, and according to the Organization of State Higher Education Executive Officers, higher education institutions in all but five states experienced real declines in per student revenues from state and local sources during the first half of the current decade, at a time of flush state and local coffers. In 2006, public four-year colleges and universities relied on tuition for over 37 percent of their total revenues for the first time in modern history, and the recent financial crisis has surely further increased the fiscal constraints faced by public and private universities alike.

The dramatically increased fiscal constraints facing public colleges and universities, coupled with rapid improvements in technology, has paved the way for higher education institutions to introduce technology-based platforms for mass instruction. The use of internet classes has exploded over the past decade, especially in the past few years. Over 2.6 million students took at least one online course in fall 2005, up from 1.6 million three years earlier (Allen and Seaman, 2006). Though the majority of these students are in community colleges and junior colleges, more than 80 percent of doctoral/research institutions in the United States offer online classes. Each of the ten largest four-year colleges and universities in the United States offers online classes, some with over 400 sections and others with more than 10,000 students per term enrolled in at

least one online class. Today, virtually every institution with more than 15,000 students offers online classes.

If internet-based classes are at least reasonable substitutes for live-lecture classes, then the use of internet-based classes could be a cost-effective method of combating increased fiscal constraint. And in theory, internet-based classes may even dominate live-lecture classes, as they offer students more flexibility in the timing of attendance as well as the opportunity to review lectures to clear up confusing points. They also provide the opportunity for improved access to higher education for residents of remote communities. On the other hand, internet-based lectures provide weaker incentives for students to regularly attend and keep up with classes, and as has been documented at one major four-year institution, last-minute cramming in internet-based courses is rampant (Donovan, Figlio and Rush, 2006). But since increasing live-lecture class sizes is associated with deleterious consequences for students (Bettinger and Long, 2007), offering classes through an electronic medium may be an appealing alternative mechanism for cost savings in higher education.

A major report released by the U.S. Department of Education on June 26, 2009 provides additional support for the expansion of online education. This study, a meta-analysis of the available research on live versus online delivery of education (primarily higher education), suggests that online delivery of material leads to improvements in student outcomes relative to live delivery, with hybrid live-plus-internet delivery having the largest benefits of all. While the Department of Education's press release on the report concentrated on the potential benefits of integrating electronic content into regular

classrooms, the ensuing news coverage (and the report itself) also emphasized the relative benefits of online-only education.

That said, the studies that provided the basis for this meta-analysis may not be sufficient to draw conclusions about the relative benefits of live versus online education. Only sixteen of the studies considered in this meta-analysis used a simple randomization method to assign students into either treatment or control groups, with an average study size of 84 participants, and only two of these studies had the same instructor teaching both the treatment and control group. In these two studies (Zhang, 2005; Zhang et al, 2006), the researcher compared a 45-minute single-session live lecture in a classroom setting to a 45-minute single-session e-learning experience in a research laboratory. Therefore, the existing experimental evidence on live versus online delivery of large lectures is effectively nonexistent.[1]

This paper aims to fill this important gap by reporting on an experiment in which students were randomly assigned to either an online or a live section of a course taught by one instructor and for which the ancillaries for the class, such as the web page, problem sets and TA support, as well as the exams, were identical between the sections. The *only*

[1] The U.S. Department of Education meta-analysis did not include a small number of papers that have been published in economics journals that we believe meet their criteria for inclusion in the federal study. Navarro and Shoemaker (2000) present evidence that students taking an introductory macroeconomic online had significantly better test scores than students taking the same course in a live lecture format, while Brown and Liedholm (2002) report just the opposite result for students taking an introductory microeconomic class. However, Navarro and Shoemaker compared a lecture format, with no class web page, to an online format which included a web page, with a bulletin board for posting questions, weekly online chat discussions with the instructor, and quizzes, which the students were required to take weekly, as well as giving the students a CD with the audio part of the lectures along with PowerPoint slides and review questions. Brown and Liedholm's online versus live comparison contrasted a live class with (apparently) no web page to an online class with streaming videos of one semester's lectures and a variety of additional material, such as numeric problems and repeatable quizzes. These are hardly "apples to apples" comparisons and so conclusions drawn from them about the performance of on-line versus live lectures are not robust.

difference between these sections is the method of delivery of the lectures: Some students viewed the lectures live, as would be the case in traditional classes, while other students viewed the lectures on the internet. Thus we are able to determine how online delivery of traditional lectures compares with live delivery. It is important to note that all students had access to a rich array of internet-based resources regardless of the location of where they viewed the lectures. In many ways, therefore, this is precisely the tradeoff that universities are increasingly facing as they decide the appropriate medium for lecture delivery in their large classes. The results of this experiment, therefore, have significant potential implications for public and university policy.

## II. THE CLASS AND THE EXPERIMENT

We utilize data from an experiment conducted in a large Principles of Microeconomics class taught at a large selective doctorate-granting university. This class is taught to between 1,600 and 2,600 students a semester by a single instructor. Typically, the students can register for a "live" section in which they can watch the lecture in a room with approximately 190 seats or they can register for an "online" section in which they watch the same lecture online. The lecture is videotaped as it is presented and then made available via the class web page to all students. Once the lecture is taped, it is retained on the Internet for the entire semester. Given the room-size constraint, most students register for an online section. In a typical semester, approximately 50 or 60 students actually come to any given live lecture. Because the room has vacant seats, normally no effort is made to keep the students who registered for an Internet section from attending the live section. In fact, because the live section is limited to 190, most of the students attending the live lecture have registered for an online section simply because the live section was

filled when it came time for them to register for the class. The majority of the students who register for the live section ultimately choose to watch the majority of the lectures online.

Students who register for the live section and students who register for the online section have access to the exact same class web page. The class web page has a link to watch the lectures as well as a substantial variety of class supplements: a set of online quizzes, past exams, and so forth. As such, both live and online students have access to a rich web-based learning environment to supplement the class lectures. The exams are given in the evening. Both sets of students take the exact same exams given at the exact same time. All students, regardless of the section for which they registered, have the same access to the instructor during office hours and have the same access to graduate student TA help. There are no discussion sessions. So in a typical semester the *only* difference between the students is the section in which they have registered, which, because anyone can attend the live lecture or watch the lecture online, is a meaningless distinction. Grading in the class is based on only exams. There are three exams: two midterms and a final exam. The exams are all multiple choice and are all machine graded. The instructor creates the exams which are primarily based on the lectures.

Because of the obvious selection problems, one cannot simply look at the difference in the performance of students who attend the live lecture versus students who watched the lectures online. So during the Spring 2007 semester, with the support of the instructor and the university, we conducted an experiment with this class. Before the class started, the instructor emailed all the students who had enrolled and offered them the chance to participate in an experiment. Of the nearly 1,600 students in the class, 327

students volunteered to be part of the experiment. The instructor promised to boost volunteers' grade by half of a letter grade at the end of the term -- the only incentive permitted by the university -- in exchange for allowing us the opportunity to randomly assign them to watching the lecture live or watching the lecture online. Students who were assigned to watch the lecture live had their class websites altered to remove access to the lecture online; otherwise no further change was made to their website. Students who were assigned to the online section were not allowed in the classroom to watch the live lecture. Indeed, for that semester only, the only students allowed in the classroom during the live lecture were students we had assigned to the live lecture or students who had registered for the live lecture and who opted to not participate in the experiment.[2] Among the 327 volunteers, 112 students were assigned to the live group and 215 were assigned to the online group. In order to start the experiment from the first day of class, the students were contacted before the add/drop deadline, which occurs a week after classes start. After their registration was completed, 15 of the 112 students assigned to the live group requested reassignment to an online session due to schedule conflicts. We made this reassignment but dropped them from the analysis, leaving a total of 97 students in the live-only section.[3]

The specific nature of participant recruitment in this experiment leads to potential statistical power and external validity issues. Institutional Review Board-imposed restrictions at the university in question made recruitment of a larger fraction of the student population into the experiment more difficult. The instructor was limited in the

[2] We stationed graduate students at the door to enforce these regulations and to compile a record of which students watched each live lecture.

[3] Later in the paper we present evidence suggesting that our results are qualitatively invariant to our decisions as to how we treat crossover cases.

degree to which he could contact the students to recruit them into the experiment, and we were limited as to the incentives that could be offered. The ideal situation from an external validity standpoint would have been to randomly assign all students to either a live or online section of the class, but this was not possible given the culture of the university, where mixed live-online classes are typically characterized by complete student autonomy. Statistical power is still not a major concern here: Even with a smaller-than-desired sample, we can still detect effects on the order of two points on a 100-point scale -- or one-fifth the size of the incentive to participate in the study. External validity issues, on the other hand, are a much bigger potential concern, as our study sample may not be representative of a broader population of potential students. We discuss the limitations to external validity in section IV below.

## III. THE DATA AND THE RESULTS

Four groups of students took the course in question:

1) Students who volunteered for the experiment and were randomly assigned to watching the lectures online. These students were required to watch the lectures online. 215 students fell within this group.

2) Students who volunteered for the experiment and were randomly assigned to watching the lectures live. These students were required to watch the lectures live. 97 students fell within this group.

3) Students who did not volunteer for the experiment and were initially registered in an online section. These students were required to watch the lectures online. 1,203 students fell within this group.

4) Students who did not volunteer for the experiment and were initially registered in the live section. These students were allowed to choose whether to watch a lecture live or online, or a hybrid thereof. 77 students fell within this group.

Groups 1 and 2, the participants in our experiment, had exactly the same course with one crucial difference: They were randomly assigned to different delivery

mechanisms for the lectures. Hence comparing their performance potentially offers us an apples-to-apples comparison of an online class to a traditional live lecture without worrying about the possibility of selection issues or how to correct for the selection.

First, however, we examine whether the students who volunteered for the experiment were different in observable ways from the non-volunteers. Table 1 compares the students who volunteered for the experiment with those who did not, for two groups of students -- those who initiated the class and those who completed the class. The data pertaining to the students' maternal educational attainment were obtained directly from the students; the remaining data were obtained from the university's records. As can be seen in the table, experiment volunteers differ from non-volunteers along a number of dimensions, but the differences are not unidirectional. For example, experiment volunteers are more likely to have higher grades at the university than are non-volunteers but volunteers tend to have lower SAT scores than do non-volunteers. In addition to these thoroughly mixed differences, the differences tend to be modest in magnitude. The SAT score difference, for example, was only 18 points -- less than 9 percent of the university's interquartile range of 220 SAT points. We therefore observe little evidence that the volunteers are markedly different than their non-volunteer classmates.

Table 2 compares the attributes of volunteers assigned to the live section versus those assigned to the online treatment. As can be seen in the table, the random assignment of volunteers successfully led to balancing of the volunteer population into the live-only section and the online-only section. Those assigned to watch the class

online had slightly higher prior university GPAs.[4] The live section also had fewer mothers who attended college but did not attain a degree, but overall average education levels are comparable across the two groups. In sum, it appears that the randomization in the experiment was largely successful in balancing the live and online delivery groups.

Table 3 presents the mean test scores for the two groups of students on each of the three examinations in the course, as well as the average of the three scores. We prefer to use the average score because it has the smallest problem with measurement error, and indeed, the standard errors are lowest with regard to the average score. Exams are scored on the standard 0 to 100 point scale, and the mean of the average score on the exams is just below 80 points. As can be seen from the table, the preponderance of the evidence indicates that students perform better in the live setting than in the online setting, though the raw differences are uneven and statistically insignificant. Students in the live section tended to do better on the first exam, the final exam, and overall while students in the online section performed trivially better on the second exam. These (basically zero) results are relatively precisely-estimated; the two-point difference in average exam scores that would be statistically detectable with the observed standard errors is small in comparison to the five-point incentive, considered modest by the university's Institutional Review Board, that was offered students to participate in the experiment. Therefore, we are confident that the statistical power issues associated with not recruiting a larger fraction of the class are not responsible for the null findings reported in Table 3.

[4] The difference is smaller for course completers than at the beginning of the course. However, there is little evidence of differential attrition; under no circumstance are the attributes of live-section attriters different from those of online-section attriters at even the 30 percent statistical significance level.

Controlling for covariates in an experimental setting can lead to improved precision as well as can help to reveal non-randomness in small samples. In the bottom row of Table 3, we therefore report the results of comparisons between live-only and internet-only groups in which we control for the covariates reported in Table 2. Controlling for covariates leads to modestly smaller standard errors and larger positive differences between the live-only and internet-only group; indeed, the positive differences are statistically significant in the case of the final exam and average scores. Therefore, the unconditional mean comparisons reported in Table 3 are likely to be understatements of the positive effects of live-only instruction relative to internet-only instruction. For the remainder of the paper, we use the more conservative unconditional mean comparisons, which we interpret as lower-bound estimates of the relative effects of live-only instruction.

We mention in footnote 3 that 15 of the 112 students initially assigned to the live-only treatment requested that they be removed from the experiment because of scheduling problems, and our results thusfar exclude these students. However, these students may be differentially selected. Therefore, in the fifth column of Table 3 we include all students who were initially assigned to the live-only group as a test to see whether these students' inclusion or exclusion is consequential. We find results in this intent-to-treat analysis that are quite similar to those found when these students are excluded from the analysis altogether.

As seen in Table 2, 16 students (6 -- or 6 percent -- in the live-only group and 10 -- or 5 percent -- in the internet-only group) left our study due to dropping the course before the final exam, and there is a smaller difference between live and internet students

in the university grade point average for course completers than for course beginners. While there are not statistically significant differences in any comparison of the attributes of those who left the live-only group versus those who left the internet-only group, suggesting that attrition appears to be largely random, we report in the final column of Table 3 the results of the first exam comparison when we limit the analysis only to course completers. Unsurprisingly, the set of course completers scored trivially better on the first exam than those who left the course, but there is no evidence of differential selection across the groups of students. Therefore, we find that the results when we limit our analysis to course completers are in the same ballpark as are those when we do not limit our analysis to course completers. As an alternative check, we conduct a bounding exercise in which we assign all attriters either a score of zero or a score of 100 on the relevant exam.[5] The results of this bounding exercise, presented in the last row of Table 3 for the three individual exams and the average exam score, suggest that no matter what we assume the test scores of attriters to be, there is little change in the results. Therefore, all available evidence suggests that attrition is not responsible for our findings.

While the overall effect of live instruction relative to internet delivery is very modest and positive (though not statistically distinguishable from zero in the unconditional mean comparisons), these mean effects may mask substantial differences in relative benefits of one medium of instruction over another. For instance, students from different language backgrounds, experience or motivation levels might have different experiences in live versus internet only settings. While we cannot directly measure these specific types of factors, we can stratify the estimated effects of live versus

[5] This is an approach used by Krueger (1999) in his analysis of the Project STAR class size experiment in Tennessee. Similar exercises have been conducted by Angrist, Bettinger and Kremer (2006) and others.

internet instruction along a few observable lines: by student race/ethnicity, sex and prior achievement levels.[6] For this last stratification, we define "high achievers" as students whose prior college GPA was greater than or equal to the median GPA and define "low achievers" as students whose prior college GPA was less than the median GPA. We report these results in Table 4. The treatment effects reported in Table 4 reflect average score differences for students enrolled in the live section versus those enrolled in the online section. We observe that for all racial/ethnic groups, for both male and female students, and for both high and low achievers, the average test score is higher for the set of students in live instruction versus those in online instruction. Importantly, in a number of cases this difference is statistically significant, and some of the estimated differences are large in magnitude. Most notably, the average test score grade for Hispanic students is dramatically higher in the case of live instruction. In addition, the estimated live instruction advantage is statistically significantly different from zero for male students and for low-achievers. (We should point out, however, that the only subgroup difference that is statistically significant, given our small sample size, is the one regarding race and ethnicity.) While it is premature to definitely ascribe a mechanism through which differential effects on outcomes may be operating, we can propose a few. For instance, perhaps low-achieving and male students are tempted to defer instruction and cram for exam in online-only classroom experiences or perhaps language-minority students have increased difficulty with listening to lectures in an internet setting. While we did not explicitly test these mechanisms, the results for the various subgroups indicates that

[6] These are the most logical ways to stratify our data given the observed background characteristics at our disposal. We were concerned that the subgroup results may be mere statistical artifacts, so we attempted a variety of stratifications of the data. In nearly every stratification we attempted, we found that at least one subgroup had statistically significantly positive estimated effects of live-only instruction.

future experimentation that paid particularly close attention to potentially sensitive student subgroups may be highly informative.

One possible threat to validity of this experiment involves the potential for contamination. While it was impossible for students not selected to be in the live section to attend the live lectures, it was certainly possible for experimental students to surreptitiously view online lectures even though they could not do so using their own accounts. Indeed, while we have no way of knowing how prevalent this behavior was, it is likely that at least some of the live-only students did this; only 32 percent of "live-only" students attended at least 90 percent of the live lectures[7], and 36 percent attended fewer than 20 percent of the live lectures! (Figure 1 presents a density plot of the live lecture attendance for the live-only students.) It is not clear whether this non-compliance would bias our estimates upward or downward. On the one hand, if the true effect of live instruction is positive, especially for some subgroups, the fact that we could not fully prevent "live-only" students from watching classes on the internet using a friend's account may mean that our results understate the true effects of live class attendance.

On the other hand, the potential contamination could upward-bias our results if our live-only treatment is really better thought of as a hybrid live-plus-internet treatment. There is, however, reason to believe that the live-only treatment is different from the traditional live-plus-internet hybrid that the 77 non-participant students registered to the live section experienced. Figure 1 compares the distribution of live lecture attendance for the live-only group versus the live-plus-internet group of students who did not participate

[7] Students can attend a maximum of 45 lectures.

in the experiment, and shows that the live-only students attended appreciably more live lectures (an average of 21.6 live lectures) than the live-plus-internet students (who averaged 13.6 live lectures.) Unfortunately, it is impossible to know how many on-line lectures were viewed by live-only students because they were officially blocked from downloading the lectures. Therefore, it is clearly the case that being officially restricted to only view lectures live strongly influenced the likelihood that the live-only participants would indeed receive their material delivery in the live format.[8] Hence, though we cannot know for certain, we suspect that contamination of our experiment due to participating students watching internet lectures is not a major force driving our findings.

It may also be the case that live-only participants benefit from having other classmates in the live section who are better or more motivated students, and who could therefore have positive peer effects. (This could happen if students who enroll in the live section are systematically higher-ability than those who enroll in the online section.) Since section registration had historically had no bearing on whether a student could attend the live lecture, we believe that it is unlikely that the non-participants in the live section would be much different from the non-participants in the online section, and indeed, this appears to be the case. In fact, if anything the non-participants in the live section have lower observables than those in the online section. For instance, when comparing students who did not participate in the experiment, we find that the mean SAT score for those in the live section is 1197 as compared with 1245 in the online section.

[8] It is also the case that many of the students who are observed rarely coming to class might actually not ever view the lectures at all. The university has several competing lecture note-taking services that are extremely popular with students. In addition, we find in our present data that students who attend fewer live lectures do substantially worse on the examinations, suggesting that many of those who attend fewer live lectures are not substituting surreptitiously downloaded internet lectures for the live lectures they eschew.

For the same two groups, maternal education is lower for live section registrants than for online section registrants (24 versus 18 percent had mothers with only a high school degree.) In summary, there exists no evidence that the live-only participants' scores are being positively influenced by an improved peer group of non-participating students who insisted on being part of the live section.

## IV. LIMITATIONS TO EXTERNAL VALIDITY AND RECOMMENDATIONS FOR FUTURE EXPERIMENTS

This paper presents the first experimental evidence of the relative efficacy of live versus internet-only instruction in a higher education setting. While our analysis has a high degree of internal validity, there are a number of key reasons why we believe that our results should be taken as suggestive rather than conclusive, and why we recommend that further experimentation in a variety of other settings take place before one can draw definitive conclusions about the effects of different modes of lecture delivery.

One reason that the external validity of our analysis is limited is that the volunteers whom we recruited may not reflect the overall population of students enrolled in the class. While our experimental live-only and internet-only groups are balanced along a large number of dimensions, participation in the experiment was voluntary. Moreover, the incentive used to induce participation was extra credit on the final course grade. There is no reason to believe that responsiveness to this incentive is exogenously-determined, and in fact, one can easily tell stories about which types of students might be willing to participate in the experiment. Specifically, one might reasonably expect that students who are motivated to achieve high grades but are relatively concerned about

their ability to earn high grades might be the students most responsive to a participation-for-points incentive, and this could explain why our volunteer participant group has slightly lower levels of SAT scores but slightly higher pre-course university grades as well as (insignificantly) higher high school grades. If people who are especially grade-motivated respond differently to live versus internet instruction, then our experiment has less to say about the typical student enrolled in a very large introductory course.

This concern yields important lessons for future experiments on this topic. While Institutional Review Board and university culture at the university in question did not permit us to randomize all students into live versus internet-only lecture categories, it will be important for future experiments to attempt to study the entire set of students who select into a given class, rather than a subsample of students. In the event that this is not possible, future experiments could improve upon the external validity of the present study by seeking to obtain a higher participant rate. The university required us to take a more passive role than would be desirable in the recruitment of students into the study; we could not, for instance, offer financial or in-kind incentives to increase participation, and we were limited in the number of times that we could attempt to recruit students. Settings with fewer such encumbrances might yield higher degrees of external validity.

Other external validity issues associated with this experiment would not be solved even had we been able to randomly assign 100 percent of the students in the introductory microeconomics class to live-only or internet-only lecture groups. One involves the specific university setting: The university in question is one where very large lecture classes are the norm for virtually all freshman and sophomore-level courses, across all

fields, and moreover, most of the core courses for students majoring in business are offered on this electronic platform. The results of an experiment in this type of university setting may not generalize to other university settings where students have less experience with large auditorium lectures and electronically-delivered lectures. Ideally, future experiments of this nature will take place in a wider variety of institutional settings, so that we can begin to understand the degree to which the findings generalize across settings. This is also the case because the university in question is one of the most selective state universities in the United States; the results may not generalize to open-enrollment institutions or those where students are drawn from lower in the ability and achievement distribution. Of course, given that our findings suggest that lower-ability students are a sub-group potentially most harmed by internet delivery, the results might be particularly relevant for less-selective institutions.

In addition, introductory economics courses are generally delivered in traditional lecture settings even at small institutions with modest class sizes. In some ways, one might expect that this would be the type of subject matter where live instruction may be the least beneficial, as members of the class tend to be relatively passive consumers of material in the lecture setting. It may be the case that live classes might be relatively more beneficial in other types of courses, with a greater role for interactive activities in the classroom. In such a case, our results might be an understatement of the effects of live versus internet class delivery in other contexts. On the other hand, introductory economics has a number of topics that build upon one another and relies more heavily on technical prerequisites than many other subjects do; it could be that the disciplined pacing that comes with live-only lectures might be relatively beneficial in this type of context,

implying that the effects in other fields where pacing is less crucial may be smaller. It is therefore important to conduct similar experiments in a wider range of subjects, and classes with different levels of student involvement and interactivity, in order to develop general conclusions about the relative efficacy of live versus internet-based instruction. Therefore, while our study represents the first causal evidence of the effects of live versus internet-based instruction in a university course delivery setting, it can only be seen as a beginning step toward understanding the generalized effects of different methods of instructional delivery.

Finally, while not a threat to external validity and generalizability, our subgroup-specific findings indicate that some student populations may be particularly sensitive to the mechanism through which lecture material is delivered. Language-minority students might have more difficulty following recorded lectures, and some students may be relatively less disciplined in keeping up with the pace of the course when procrastination is more possible. (This might be an explanation for the relatively large estimated effects observed for male and lower-achievement students, though we cannot say for certain that this is the reason.) Therefore, future experimentation that could directly test for some of these potential mechanisms could be highly valuable. For example, if one is interested in seeing whether delayed lecture viewing is a potential mechanism generating lower outcomes for internet-only students, one might design an experiment in which students were required to download (or maybe view) lectures within a certain number of days following lecture recording. In general, it would be highly valuable to look more deeply at the potential causal mechanisms through which different lecture delivery mechanisms might affect student learning. Additional survey and qualitative work on questions such

as ways in which students engage with the course material, interact with the instructors and their peers, pay attention to lectures and study for examinations would be highly valuable, and could help universities and professors refine their courses and instructional delivery to maximize student learning.

## V. CONCLUSION

Given the clear scale economies associated with online instruction, educational institutions are actively incorporating online instruction into their portfolios. While our results are not definitive, and numerous potential external validity concerns exist, our experimental, apples-to-apples comparison nonetheless indicates that a rush to online education may come at more of a cost than educators may suspect.

It may still be the case that the educational costs associated with internet-only classes are outweighed by the economies of scale associated with offering traditional classes over the internet. A benefit-cost analysis of internet-based lecture delivery would weigh the educational costs of reduced human capital against the financial benefit of exploiting scale economies.

On the financial benefit side, universities might potentially enjoy tremendous financial savings by recording lectures and replaying them for multiple years with an instructor taking a less hands-on approach. But even assuming courses could be devised and delivered anew each term, there may be financial advantages to offering internet delivery of courses. The university where the experiment was conducted currently administers some completely internet-only distance-education courses in addition to the hybrid live-internet courses like the one that formed the basis for this experiment. At this

university, the average administrative expenditure (not counting instructor compensation) per student per class was about $270 in 2009-10. These calculations include technical staffing, assessment administration, camera operation, and other expenditures. None of these distance classes are offered at the ideal scale for this thought experiment, but if a fully-scaled-up course of 1,000 students would cost half that at $135,000 to deliver (a reasonable estimate based on the staffing and technical scenarios that the university describes) then whether this is a large or small number would depend on the costs of staffing live courses versus internet courses. If five 200-student live lectures represent 1.5 faculty positions and a 1,000 student internet lecture course represents one-half of a faculty position[9], then under these assumptions an internet-only course would be cost-effective -- assuming there is no cost to reduced human capital -- for faculty salaries (including benefits) exceeding $135,000 (translating to a salary of about $111,600 given the university's fringe benefit rate.)[10] According to American Association of University Professors faculty salary survey data for four-year colleges and universities in 2010-11, in only three of 31 broad fields is the average *full professor* salary greater than this value, and the field with the highest average salaries for instructors (legal professions and studies) averaged $64,785.[11] Therefore, in order for internet-based delivery of traditional lectures to be financially beneficial, a college or university would typically need to provide fewer support services for internet-based classes, or use the same lectures for a number of years with a lower level of faculty participation.

[9] We assume that a university would either need to count a very large course as more than a regular course or to compensate the faculty member for the extra burden of administering a course with many students. The university in question does the latter, and compensates the faculty member an extra $50 per student in distance-education courses.

[10] This is likely a lower bound of this critical value, as there would be start-up costs associated with investment in the infrastructure at the institution.

[11] The critical value is lower than the average salary of economics faculty members at research universities, though well above the average salary of economics instructors at these universities.

These calculations make clear that, setting aside educational consequences, internet-based courses can only make financial sense either at universities with low faculty teaching loads or fields with relatively high faculty salaries, *or* at other institutions/fields if a course is developed once and replayed with much lower levels of faculty involvement. The results of this experiment suggest that the break-even point at which internet-only lecture-based classes pass a benefit-cost test is higher than the figures described above. How much higher this threshold would be depends on the costs to students associated with diminished human capital. And the threshold would be higher still if an internet-based course with fewer student services and less professor contact would yield worse results than the high-contact internet-based course described in this study. The degree to which this is true is not yet known.

At the least, our findings indicate that much more experimentation is necessary before one can credibly declare that online education is peer to traditional live classroom instruction, let alone superior to live instruction. While online instruction may be more economical in some circumstances to deliver than live instruction, our results indicate that -- consistent with a fundamental lesson of principles of microeconomics -- the lunch may be less free than many might believe.

Table 1: Baseline Summary Statistics: Volunteers versus Non-Volunteers

| | Students beginning the semester | | | Students ending the semester | | |
|---|---|---|---|---|---|---|
| Variable | Volunteer | Non-volunteer | Difference | Volunteer | Non-volunteer | Difference |
| Number of observations | 312 | 1286 | | 296 | 1186 | |
| University GPA | 3.262 | 3.156 | 0.106** | 3.282 | 3.198 | 0.084* |
| | (0.036) | (0.022) | (0.048) | (0.036) | (0.022) | (0.048) |
| SAT score | 1224.589 | 1242.884 | -18.295* | 1228.864 | 1251.541 | -20.108* |
| | (8.774) | (4.77) | (10.623) | (8.826) | (4.823) | (10.31) |
| ACT score | 25.776 | 26.482 | -0.706 | 25.921 | 26.648 | -0.727 |
| | (0.353) | (0.233) | (0.487) | (0.364) | (0.241) | (0.5) |
| High school GPA | 3.743 | 3.649 | 0.094 | 3.762 | 3.688 | 0.074 |
| | (0.053) | (0.03) | (0.067) | (0.054) | (0.031) | (0.067) |
| Female | 0.546 | 0.496 | 0.05 | 0.541 | 0.487 | 0.053 |
| | (0.028) | (0.014) | (0.032) | (0.029) | (0.015) | (0.032) |
| Black | 0.115 | 0.087 | 0.028 | 0.108 | 0.074 | 0.034* |
| | (0.018) | (0.008) | (0.018) | (0.018) | (0.008) | (0.018) |
| White | 0.610 | 0.638 | -0.028 | 0.622 | 0.648 | -0.026 |
| | (0.028) | (0.013) | (0.03) | (0.028) | (0.014) | (0.031) |
| Asian | 0.109 | 0.098 | 0.011 | 0.108 | 0.102 | 0.006 |
| | (0.018) | (0.008) | (0.019) | (0.018) | (0.009) | (0.02) |
| Hispanic | 0.115 | 0.135 | -0.02 | 0.111 | 0.136 | -0.024 |
| | (0.018) | (0.01) | (0.021) | (0.018) | (0.01) | (0.022) |
| Mother attended high school only | 0.193 | 0.181 | 0.011 | 0.193 | 0.181 | 0.011 |
| | (0.023) | (0.011) | (0.025) | (0.023) | (0.011) | (0.025) |
| Mother attended some college | 0.176 | 0.179 | -0.004 | 0.176 | 0.179 | -0.004 |
| | (0.022) | (0.011) | (0.025) | (0.022) | (0.011) | (0.025) |
| Mother graduated from college | 0.301 | 0.401 | -0.100*** | 0.301 | 0.401 | -0.100*** |
| | (0.027) | (0.015) | (0.032) | (0.027) | (0.015) | (0.032) |
| Mother earned graduate degree | 0.223 | 0.195 | 0.028 | 0.223 | 0.195 | 0.028 |
| | (0.024) | (0.012) | (0.026) | (0.024) | (0.012) | (0.026) |
| Mother's education unknown | 0.054 | 0.042 | 0.012 | 0.054 | 0.042 | 0.012 |
| | (0.057) | (0.029) | (0.011) | (0.057) | (0.029) | (0.011) |

Notes: Standard errors are presented in parentheses beneath means. Differences marked ***, ** and * are statistically significant at the 1, 5 and 10 percent levels, respectively. Sixteen volunteers and 100 non-volunteers dropped the course during the semester. Questions regarding maternal education were asked during the final examination, so we only have these variables for students who completed the course; therefore, the first and second sets of columns are identical for these variables.

Table 2: Baseline Summary Statistics: Volunteers Assigned to Live Section versus Volunteers Assigned to Online Section

| | Students beginning the semester | | | Students ending the semester | | |
|---|---|---|---|---|---|---|
| Variables | Live | Online | Difference | Live | Online | Difference |
| Number of observations | 97 | 215 | | 91 | 205 | |
| University GPA | 3.138 | 3.321 | -0.183 | 3.193 | 3.321 | -0.128 |
| | (0.073) | (0.04) | (0.077)** | (0.071) | (0.042) | (0.079) |
| SAT score | 1214.133 | 1228.581 | -14.448 | 1216.447 | 1228.581 | -12.134 |
| | (15.944) | (10.519) | (18.751) | (15.902) | (10.519) | (18.697) |
| ACT score | 25.75 | 25.784 | -0.034 | 26 | 25.898 | 0.102 |
| | (0.602) | (0.426) | (0.833) | (0.629) | (0.435) | (0.882) |
| High School GPA | 3.794 | 3.718 | 0.076 | 3.847 | 3.724 | 0.123 |
| | (0.08) | (0.068) | (0.115) | (0.073) | (0.071) | (0.118) |
| Female | 0.495 | 0.567 | -0.072 | 0.473 | 0.571 | -0.098 |
| | (0.051) | (0.034) | (0.061) | (0.053) | (0.035) | (0.063) |
| Black | 0.124 | 0.112 | 0.012 | 0.121 | 0.102 | 0.018 |
| | (0.034) | (0.022) | (0.039) | (0.034) | (0.021) | (0.039) |
| Asian | 0.093 | 0.116 | -0.023 | 0.099 | 0.112 | -0.013 |
| | (0.03) | (0.022) | (0.038) | (0.031) | (0.022) | (0.039) |
| White | 0.639 | 0.600 | 0.039 | 0.637 | 0.615 | 0.023 |
| | (0.049) | (0.033) | (0.060) | (0.051) | (0.034) | (0.061) |
| Hispanic | 0.082 | 0.126 | -0.044 | 0.088 | 0.122 | -0.034 |
| | (0.028) | (0.023) | (0.039) | (0.03) | (0.023) | (0.04) |
| Mother attended high school only | 0.196 | 0.177 | 0.019 | 0.209 | 0.185 | 0.023 |
| | (0.041) | (0.026) | (0.047) | (0.027) | (0.027) | (0.050) |
| Mother attended some college | 0.093 | 0.200 | -0.107 | 0.099 | 0.210 | -0.111 |
| | (0.03) | (0.027) | (0.045)** | (0.029) | (0.029) | (0.048)** |
| Mother graduated from college | 0.309 | 0.274 | 0.035 | 0.330 | 0.288 | 0.042 |
| | (0.047) | (0.031) | (0.055) | (0.032) | (0.032) | (0.058) |
| Mother earned graduate degree | 0.227 | 0.205 | 0.022 | 0.242 | 0.215 | 0.027 |
| | (0.043) | (0.028) | (0.050) | (0.029) | (0.029) | (0.053) |
| Mother's education unknown | 0.072 | 0.042 | 0.030 | 0.077 | 0.044 | 0.033 |
| | (0.026) | (0.014) | (0.027) | (0.014) | (0.014) | (0.029) |

Notes: Standard errors are presented in parentheses beneath means. Differences marked ***, ** and * are statistically significant at the 1, 5 and 10 percent levels, respectively.

Table 3: Comparison of Average Test Scores for Live Versus Online Instruction

| Section | Exam one | Exam two | Final exam | Average score | Average score-treating defectors as "live" | Exam one-excluding attrition |
|---|---|---|---|---|---|---|
| Number of observations | 312 | 301 | 296 | 296 | 311 | 296 |
| Live | 84.536<br>(1.168)<br>[97] | 76.692<br>(1.193)<br>[93] | 75.939<br>(0.837)<br>[91] | 79.940<br>(0.85)<br>[91] | 79.948<br>(0.876)<br>[106] | 85.925<br>(1.093)<br>[91] |
| Online | 83.301<br>(0.957)<br>[215] | 76.904<br>(0.876)<br>[208] | 74.302<br>(0.799)<br>[205] | 78.502<br>(0.675)<br>[205] | 78.571<br>(0.711)<br>[205] | 84.137<br>(0.943)<br>[205] |
| Difference | 1.235<br>(1.626) | −0.212<br>(1.534) | 1.637<br>(1.426) | 1.440<br>(1.209) | 1.377<br>(1.179) | 1.788<br>(1.592) |
| Difference conditional on covariates | 1.981<br>(1.562) | 2.270<br>(1.466) | 3.000<br>(1.409)** | 2.508<br>( 1.130)** | 2.230<br>(1.112)** | 2.017<br>(1.542) |
| Upper/lower bound estimates | [1.94, 2.03] | [2.20, 2.39] | [2.86, 3.12] | [2.37, 2.61] | | |

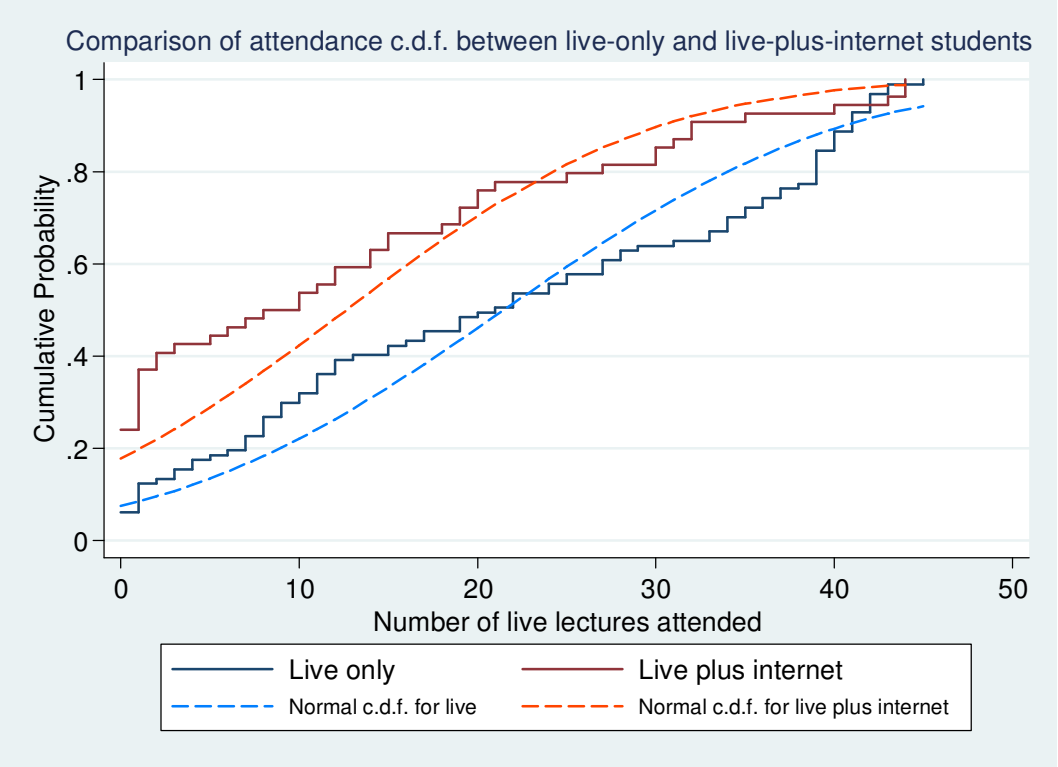
Notes: Standard errors are in parentheses beneath coefficient estimates. The dependent variable is the exam score measured on a 0-to-100 point scale. Differences marked ***, ** and * are statistically significant at the 1, 5 and 10 percent levels, respectively.

Table 4: Heterogeneous Effects of Live Instruction Versus Online Instruction

| Subgroup | Results by racial/ethnic group | Results by student sex | Results by achievement level |
|---|---|---|---|
| White students | 1.117 | | |
| | (1.436) | | |
| Black students | 2.828 | | |
| | (3.239) | | |
| *P-value* | [0.632] | | |
| Hispanic students | 11.276*** | | |
| | (3.587) | | |
| *P-value* | [0.014] | | |
| Asian students | 4.319 | | |
| | (3.590) | | |
| *P-value* | [0.428] | | |
| Male students | | 3.480** | |
| | | (1.680) | |
| Female students | | 1.780 | |
| | | (1.576) | |
| *P-value* | | [0.462] | |
| Low-achievers | | | 4.054*** |
| | | | (1.536) |
| High-achievers | | | 1.169 |
| | | | (1.635) |
| *P-value* | | | [0.205] |
| R-squared | 0.386 | 0.370 | 0.402 |

Notes: Dependent variable is the average test score measured on a 0-to-100 point scale. Number of observations: 296. Standard errors are in parentheses beneath coefficient estimates. Differences marked ***, ** and * are statistically significant at the 1, 5 and 10 percent levels, respectively.

Figure 1. End of semester attendance cumulative density functions for live-only (lower lines) and live-plus-internet (upper lines) students



REFERENCES

Allen, I. Elaine, and Jeff Seaman. 2006. *Making the grade: Online education in the United States, 2006.* Boston, MA: Sloan Consortium.

Angrist, Josh, Eric Bettinger, and Michael Kremer. 2006. Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia." *American Economic Review* 96, no. 3:847-862.

Bettinger, Eric, and Bridget T. Long. 2007. Institutional responses to reduce inequalities in college outcomes: Remedial and developmental courses in higher education. In *Economic inequality and higher education: Access, persistence, and success*, eds. Stacy Dickert-Conlin and Ross Rubenstein. New York: Russell Sage Foundation.

Brown, Byron W., and Carl E. Liedholm. 2002. Can web courses replace the classroom in principles of microeconomics? *American Economic Review* 92, no. 2:444-448.

Donovan, Colleen, David Figlio, and Mark Rush. 2006. Cramming: The effects of school accountability on college-bound students. Working paper, National Bureau of Economic Research, Cambridge, MA.

Kane, Thomas, and Peter Orszag. 2003. Funding restrictions at public universities: Effects and policy implications." Working paper, Brookings Institution, Washington, DC.

Krueger, Alan. 1999. Experimental estimates of education production functions. *Quarterly Journal of Economics* 114, no. 2:497-532.

Mortenson, Thomas. 2005. State tax fund appropriations for higher education: FY 1961 to FY 2005. Report, Postsecondary Education Opportunity.

Navarro, Peter, and Judy Shoemaker. 2000. Policy issues in the teaching of economics in cyberspace: research design, course design, and research results. *Contemporary Economic Policy* 18, no. 3:359-366.

U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. 2009. *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Washington, DC: U.S. Department of Education.

Zhang, Dongsong. 2005. Interactive multimedia-based e-learning: A study of effectiveness. *American Journal of Distance Education* 19, no. 3:149–62.

Zhang, Dongsong, Lina Zhou, Robert O. Briggs, and Jay F. Nunamaker, Jr. 2006. Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information and Management* 43, no. 1:15–27.