

NBER WORKING PAPER SERIES

IS IT LIVE OR IS IT INTERNET? EXPERIMENTAL ESTIMATES OF THE EFFECTS
OF ONLINE INSTRUCTION ON STUDENT LEARNING

David N. Figlio
Mark Rush
Lu Yin

Working Paper 16089
<http://www.nber.org/papers/w16089>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2010

We appreciate the financial support of the National Science Foundation and the U.S. Department of Education (via the National Center for the Analysis of Longitudinal Data in Education Research.)

We are grateful to the introductory microeconomics professor and the large university that permitted us to randomly assign students to live and online versions of the same class. All errors are our own, and our results and conclusions do not necessarily reflect the views of the National Science Foundation, U.S. Department of Education, or the undisclosed university in question. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2010 by David N. Figlio, Mark Rush, and Lu Yin. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

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

David N. Figlio, Mark Rush, and Lu Yin

NBER Working Paper No. 16089

June 2010

JEL No. I20,I23

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. Counter to the conclusions drawn by a recent U.S. Department of Education meta-analysis of non-experimental analyses of internet instruction in higher education, 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.

David N. Figlio
Institute for Policy Research
Northwestern University
2040 Sheridan Road
Evanston, IL 60208
and NBER
figlio@northwestern.edu

Lu Yin
University of Florida
Department of Economics
P.O. Box 117140
Gainesville, FL 32611-7140
luyin@ufl.edu

Mark Rush
University of Florida
Department of Economics
P.O. Box 117140
Gainesville, FL 32611-7140
mark.rush@cba.ufl.edu

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 very 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. 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. While the authors of the meta-analysis identified more than one thousand studies of online learning, they found only 51 studies that were at least "quasi-experimental," which they defined as including control variables in a cross-sectional setting. Of these 51 studies, only 28 studies directly compared an online learning condition with a face-to-face control condition. Sixteen of these studies used a simple randomization method to assign students into either treatment or control groups, with an average study size of 84 participants. But only two of these studies had the same instructor teaching both the treatment and control group, and none of these studies further controlled for measured student background characteristics, a potential problem especially in small sample situations. In the remaining two papers (Zhang, 2005; Zhang et al, 2006), written by the same author, 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. This is very different from a full-term course. In summary, *none* of the studies cited in the widely-publicized meta-analysis released by the U.S. Department of Education included randomly-assigned students taking a full-term course, with live versus online delivery mechanisms, in settings that could be directly compared (i.e., similar instructional materials delivered by the same instructor.) The evidence base on the relative benefits of live versus online education is therefore tenuous at best.¹

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

From a public and university policy standpoint, the current state of this research is dismaying: More students are being exposed to internet classes yet there is no satisfactory research demonstrating whether such changes help, hinder, or have no effect on student learning. 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* 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 lectures compares with live delivery.

The results of this experiment, therefore, have significant implications for public and university policy: If our results suggest that internet delivery of classes is inferior to live delivery, then for classes in which lectures are an important component, to make the online delivery of these classes comparable from the student-learning perspective will require generally elaborate web pages and other means of instruction and even then may be insufficient. However, if our results suggest that internet delivery of these classes is comparable to or more favorable than live delivery, then colleges and universities might be emboldened to move even more classes to the internet – though the comparison should

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.

be between large-lecture introductory courses and their internet counterparts.² However, it is important to note that the results of any one study, even one with high degrees of internal validity, should be treated with caution. Section IV of this paper highlights the limitations to generalizability of this paper, and makes recommendations for future experiments on the topic.

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 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

² Our data are from a large introductory class and so our results probably should not be applied to courses where material is delivered in smaller sections. However, it is generally the large classes that college administrations are most eager to move online.

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, 327³ students volunteered to be part of the experiment. If they volunteered, the instructor promised to boost their grade by half of a letter grade at the end of the term. In exchange, they allowed 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. We stationed graduate students at the door to enforce these regulations and to compile a record of which students watched each live lecture.

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 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.⁴ The ideal situation from an external validity standpoint would have been to randomly assign all students to either a

³ Among the 327 volunteers, 112 students were assigned to the live group and 215 assigned to the online group. In order to start the experiment from the first day of the 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 section due to schedule conflicts. We made this reassignment but then dropped them from the analysis, leaving a total of 97 students in the live-only section.

⁴ Specifically, the only incentive we were allowed to offer was a five point boost in students' test scores.

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. Of the two potential concerns -- statistical power and external validity -- associated with the recruitment of the experimental sample, external validity is more important. While we would have liked to have recruited a larger study sample, our sample size is nonetheless large enough to detect modest estimated effects of the treatment. Specifically, we can detect effects on the order of two points on a 100-point scale -- or 40 percent 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.

It is important to note that the course is already a hybrid between lectures and a rich set of internet-based applications. Therefore, one cannot view this experiment as comparing between two purely lecture-based mediums of delivery. Rather, the appropriate comparison is between two cases in which there exists considerable internet-based material, but in one case the lecture portion is delivered live and paced throughout the term and in the other case the lecture portion is delivered electronically and downloaded on demand by the student. 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.

III. THE DATA AND THE RESULTS

We have four groups of students:

- 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.⁵
- 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.

Of these four groups, most interest focuses on comparing Groups 1 and 2. These students 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

⁵ Fewer students were assigned to the live lecture than the Internet lectures because of the seating capacity constraint imposed by the lecture room.

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. Volunteers' mothers were less likely to have graduated from college but, though statistically insignificant, volunteer's mothers were more likely to have earned a graduate degree. 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 out of averages exceeding 1,200. So, at least for observable measures, we see no compelling evidence that the volunteers are markedly different than their non-volunteer classmates.

More directly relevant to the experiment is whether the attributes of volunteers assigned to the live section are comparable to those of volunteers assigned to the online section. These differences are reported in Table 2. 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. In the initial assignment, those watching the class online had slightly higher prior university GPAs, but that difference disappeared by the end of the class. The live section also had fewer mothers who attended college but then apparently dropped out. However the live section also had fewer mothers attending college at all, so any difference in family educational attainment must be slim.

Because we see no significant, consistent differences between our volunteer groups, we can proceed to examine their relative performances on the exams. 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 modestly better on the second exam (though the magnitude of the difference is small).⁶ It is evident from the simple means comparisons that students in the online only setting did not perform better than did the students in the live only setting, a finding at odds with the conclusion of the U.S. Department of Education report on internet-based education. Moreover, 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 that was offered students to participate in the experiment. Given that the university's Institutional Review Board deemed the five-point incentive to be of modest magnitude, by definition the realized differences in test scores are even more modest in magnitude. 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.

⁶ The difference in the withdrawal rate was small: Approximately 6 percent of the live section withdrew over the semester while slightly more than 4.5 percent of the online section withdrew.

While the overall effect of live instruction relative to internet delivery is very modest and positive (though not statistically distinguishable from zero), 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 internet instruction along a few observable lines: by student race/ethnicity, sex and prior achievement levels.⁷ 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. While it is premature to definitely ascribe a mechanism through which this may be operating, we can

⁷ These are the most logical ways to stratify our data given the limited number of 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.

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 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, 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, and 36 percent attended fewer than 20 percent of the live lectures! 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 some or all 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; the typical live-only participant in the experiment attended 70 percent more lectures than the typical live lecture non-participant, and was more than three times as likely as the typical live lecture non-participant to attend at least 90 percent of the lectures. Obviously many of the live lecture non-participant students opted to view the lectures on-line. 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 better 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, the mean SAT score for non-participants in the live section is 1197 while the mean SAT score for non-participants in the online section is 1245. While 24 percent of non-participants in the

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. This could explain why Donovan et al. (2006) demonstrate that a significant number of students in large internet-and-live classes at a major selective state university rarely view lectures in any form. 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.

live-only section had mothers with only a high school degree, the corresponding value for online-only non-participants is 18 percent. 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. In this section, we document some of the limitations to external validity, and we offer recommendations for future experiments regarding the relative efficacy of different lecture delivery mechanisms.

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 (and also higher high school grades, though the high school grades difference is not statistically significant.) If people who are especially motivated to attain high grades 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, regardless of whether they wished to participate in a randomized experiment, 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. We were limited by the university's Institutional Review Board 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 contact students to encourage them to participate. Future experiments in settings that have fewer such encumbrances might have higher degrees of external validity.

There are other external validity issues associated with this experiment that could 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 schools where more lower-ability students are educated. These results indicate that less selective schools should not rush to internet delivery of lectures, and instead should experiment with the relative benefits of live versus internet courses.

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 clear, therefore, that it is important to conduct similar experiments in a wider range of subject matter classes, and classes with different levels of student direct involvement and interactivity, in order to develop general conclusions about the relative efficacy of live versus internet-based instruction.

In summary, 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. In order to know more generally whether live lectures dominate internet-based lectures, under what circumstances, and for whom, numerous additional experiments will need to be conducted, in a variety of institutional settings, in different class sizes, and in different subject areas.

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. The recent U.S. Department of Education report suggesting that online instruction may be more

effective as well as being more efficient seems likely to only accelerate this process. But the papers surveyed in that meta-analysis are not experimental in general, and those that are experimental do not make apples-to-apples comparisons. The results of our experimental, apples-to-apples comparison indicate that a rush to online education may come at more of a cost than educators may suspect, given the Department of Education's report.

We do not claim that our results are definitive. Our experiment was only conducted one time, in a large course with significant internet resources available already for students taking live-instruction classes. We were not able to randomly assign all students to live versus internet delivery settings, and were forced to rely on voluntary participation in the experiment, so while internal validity is high, the results may not generalize to the student population as a whole. Furthermore, the institution is a major, very selective university. That said, our strongest findings in favor of live instruction are for the relatively low-achieving students, male students, and Hispanic students. These are precisely the students who are more likely to populate the less selective universities and community colleges. These students may well be disadvantaged by the movement to online education and, to the extent that it is the less selective institutions and community colleges that are most fully embracing online education, inadvertently they may be harming a significant portion of their student body.

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 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

Variable	Students beginning the semester			Students ending the semester		
	Volunteer	Non-volunteer	Difference	Volunteer	Non-volunteer	Difference
Number of observations	312	1286		296	1186	
University GPA	3.262 (0.036)	3.156 (0.022)	0.106** (0.048)	3.282 (0.036)	3.198 (0.022)	0.084* (0.048)
SAT score	1224.589 (8.774)	1242.884 (4.77)	-18.295* (10.623)	1228.864 (8.826)	1251.541 (4.823)	-20.108* (10.31)
ACT score	25.776 (0.353)	26.482 (0.233)	-0.706 (0.487)	25.921 (0.364)	26.648 (0.241)	-0.727 (0.5)
High school GPA	3.743 (0.053)	3.649 (0.03)	0.094 (0.067)	3.762 (0.054)	3.688 (0.031)	0.074 (0.067)
Female	0.546 (0.028)	0.496 (0.014)	0.05 (0.032)	0.541 (0.029)	0.487 (0.015)	0.053 (0.032)
Black	0.115 (0.018)	0.087 (0.008)	0.028 (0.018)	0.108 (0.018)	0.074 (0.008)	0.034* (0.018)
White	0.61 (0.028)	0.638 (0.013)	-0.028 (0.03)	0.622 (0.028)	0.648 (0.014)	-0.026 (0.031)
Asian	0.109 (0.018)	0.098 (0.008)	0.011 (0.019)	0.108 (0.018)	0.102 (0.009)	0.006 (0.02)
Hispanic	0.115 (0.018)	0.135 (0.01)	-0.02 (0.021)	0.111 (0.018)	0.136 (0.01)	-0.024 (0.022)
Mother attended high school only	0.193 (0.023)	0.181 (0.011)	0.011 (0.025)	0.193 (0.023)	0.181 (0.011)	0.011 (0.025)
Mother attended some college	0.176 (0.022)	0.179 (0.011)	-0.004 (0.025)	0.176 (0.022)	0.179 (0.011)	-0.004 (0.025)
Mother graduated from college	0.301 (0.027)	0.401 (0.015)	-0.100*** (0.032)	0.301 (0.027)	0.401 (0.015)	-0.100*** (0.032)
Mother earned graduate degree	0.223 (0.024)	0.195 (0.012)	0.028 (0.026)	0.223 (0.024)	0.195 (0.012)	0.028 (0.026)
Mother's education unknown	0.054 (0.057)	0.042 (0.029)	0.012 (0.011)	0.054 (0.057)	0.042 (0.029)	0.012 (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

Variables	Students beginning the semester			Students ending the semester		
	Live	Online	Difference	Live	Online	Difference
Number of observations	97	215		91	205	
University GPA	3.138 (0.073)	3.321 (0.04)	-0.183 (0.077)**	3.193 (0.071)	3.321 (0.042)	-0.128 (0.079)
SAT score	1214.133 (15.944)	1228.581 (10.519)	-14.448 (18.751)	1216.447 (15.902)	1228.581 (10.519)	-12.134 (18.697)
ACT score	25.75 (0.602)	25.784 (0.426)	-0.034 (0.833)	26 (0.629)	25.898 (0.435)	0.102 (0.882)
High School GPA	3.794 (0.08)	3.718 (0.068)	0.076 (0.115)	3.847 (0.073)	3.724 (0.071)	0.123 (0.118)
Female	0.495 (0.051)	0.567 (0.034)	-0.072 (0.061)	0.473 (0.053)	0.571 (0.035)	-0.098 (0.063)
Black	0.124 (0.034)	0.112 (0.022)	0.012 (0.039)	0.121 (0.034)	0.102 (0.021)	0.018 (0.039)
Asian	0.093 (0.03)	0.116 (0.022)	-0.023 (0.038)	0.099 (0.031)	0.112 (0.022)	-0.013 (0.039)
White	0.639 (0.049)	0.6 (0.033)	0.039 (0.060)	0.637 (0.051)	0.615 (0.034)	0.023 (0.061)
Hispanic	0.082 (0.028)	0.126 (0.023)	-0.044 (0.039)	0.088 (0.03)	0.122 (0.023)	-0.034 (0.04)
Mother attended high school only	0.196 (0.041)	0.177 (0.026)	0.019 (0.047)	0.209 (0.027)	0.185 (0.027)	0.023 (0.050)
Mother attended some college	0.093 (0.03)	0.2 (0.027)	-0.107 (0.045)**	0.099 (0.029)	0.21 (0.029)	-0.111 (0.048)**
Mother graduated from college	0.309 (0.047)	0.274 (0.031)	0.035 (0.055)	0.330 (0.032)	0.288 (0.032)	0.042 (0.058)
Mother earned graduate degree	0.227 (0.043)	0.205 (0.028)	0.022 (0.050)	0.242 (0.029)	0.215 (0.029)	0.027 (0.053)
Mother's education unknown	0.072 (0.026)	0.042 (0.014)	0.03 (0.027)	0.077 (0.014)	0.044 (0.014)	0.033 (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
Number of observations	312	301	296	296
Live	84.536 (1.168) [97]	76.692 (1.193) [93]	75.939 (0.837) [91]	79.940 (0.850) [91]
Online	83.301 (0.957) [215]	76.904 (0.876) [208]	74.302 (0.799) [205]	78.502 (0.675) [205]
Difference	1.235 (1.626)	-0.212 (1.534)	1.637 (1.426)	1.440 (1.209)

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)		
Hispanic students	11.276*** (3.587)		
Asian students	4.319 (3.590)		
Male students		3.480** (1.680)	
Female students		1.780 (1.576)	
Low-achievers			4.054*** (1.536)
High-achievers			1.169 (1.635)
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.

REFERENCES

- Allen, I. Elaine and Jeff Seaman. *Making the Grade Online Education in the United States, 2006*. Massachusetts, Sloan Consortium, 2006.
- Bettinger, Eric and B. T. Long. "Institutional Responses to Reduce Inequalities in College Outcomes: Remedial and Developmental Courses in Higher Education." In S. Dickert-Conlin & R. Rubenstein (Eds.). *Economic inequality and higher education: Access, persistence, and success*. New York: Russell Sage Foundation. 2007
- Brown, Byron W. and Carl E. Liedholm. "Can Web Courses Replace the Classroom in Principles of Microeconomics?" *American Economic Review*, May 2002 (*Papers and Proceedings*), 92(2), pp. 444-448.
- Donovan, Colleen, David Figlio, and Mark Rush. "Cramming: The Effects of School Accountability on College-Bound Students." Working paper, National Bureau of Economic Research, October. 2006.
- Kane, Thomas and Peter Orszag. "Funding Restrictions at Public Universities: Effects and Policy Implications." Working paper, Brookings Institution, September, 2003.
- Mortenson, T. "State Tax Fund Appropriations for Higher Education: FY 1961 to FY 2005." Postsecondary Education Opportunity, January 2005.
- Navarro, Peter and Judy Shoemaker. "Policy Issues in the Teaching of Economics in Cyberspace: Research Design, Course Design, and Research Results." *Contemporary Economic Policy*, July 2000, 18(3), pp. 359-366.
- U.S. Department of Education, Office of Planning, Evaluation, and Policy Development, *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies*, Washington, D.C., 2009.
- Zhang, D. Interactive multimedia-based e-learning: A study of effectiveness. *American Journal of Distance Education*, 2005, 19(3), pp 149–62.
- Zhang, D., L. Zhou, R. O. Briggs, and J. F. Nunamaker, Jr. Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information and Management*, 2006, 43(1), pp15–27.