Online Vs. Face-to-Face: A Comparison of Student Outcomes with Random Assignment

J. J. Arias
Department of Economic and Finance
Georgia College & State University
Milledgeville, GA 31061
Email: jj.arias@gcsu.edu

John Swinton
Department of Economic and Finance
Georgia College & State University
Milledgeville, GA 31061

Kay Anderson
Registrar's Office
Georgia College & State University
Milledgeville, GA 31061

Abstract

The following study contrasts the efficacy of online delivery relative to face-to-face delivery using an enrolment protocol that largely eliminates self-selection bias. Only a few previous studies even attempt to control for sample selection. The study utilizes random assignment of the registrants of a Principles of Macroeconomics class into two alternative venues: online and face-to-face. The same professor taught both sections with the same course objectives and exams. Both the change in student scores from the pre-test to the post-test and the student's exam average are modelled as a function of the course environment, the student's SAT math score (or ACT equivalent), the student's GPA prior to taking the course, the student's gender and the student's overall credit hours prior to taking the course. The pre- and post-test had both standardized and instructor-specific questions. Students in the face-to-face section have statistically significantly higher exam scores and statistically significantly greater improvement on the post-test instructor questions. There is no statistical difference in the improvement on the post-test overall nor in the improvement in the post-test standardized questions. These mixed results suggest that both course objectives and the mechanism used to assess the relative effectiveness of the two modes of education may play an important part in determining the relative effectiveness of alternative delivery methods.

Keywords: Online education; e-learning; face-to-face teaching; economics.

Introduction

Online educational opportunities have blossomed as parents, students, college and university administrators and state and federal legislatures try to grapple with the problem of increasing education costs. The potential advantages of offering courses online are numerous: There is a perception that online classes are a more cost-effective way to offer some courses. Students and teachers need not physically meet in a classroom. Therefore, people in remote areas can have access to courses to which they might not have had access otherwise. In the case of asynchronous courses, students can more easily fit their learning time into their schedule. This allows more flexibility, particularly to the non-traditional students who may have family or work obligations not normally associated with the traditional undergraduate student population.   More students can consume the material simultaneously without stretching classroom capacity.

At the same time, according to a 2016 study by the World Economic Forum (World Economic Forum, 2016), the demands of the 21st Century workplace increasingly require students to master a more extensive set of skills, such as collaboration and problem-solving, than past generations had to learn. A college degree in and of itself is not as important as the mastery of needed skills for many employers (Calderon and Sidhu, 2014). Herein lays the conundrum: Does the online classroom represent a reasonable substitute for the traditional face-to-face classroom? Are the skills that students master comparable between the two delivery approaches?

For all of the advantages online classes offer, doubts remain as to whether or not online education can live up to its promises. For example, Hoxby (2014) examines the sustainability of online education at both non-selective and highly selective institutions. She concludes that the massive use of online education is only sustainable with some non-selective institutions. In a separate study, Hoxby (2017) also finds that there is little to no evidence of either large cost savings or large returns-on-investment for online education. (In fact, she finds that students personally pay more for online education relative to face-to-face education.) Although the online approach offers freedom, it requires more discipline from both students and educators. Students must make the effort to complete the material within the required time frame. They need to muster the discipline to progress through the class in a timely manner – a discipline traditionally imposed by the class schedule. When a class does not meet in a particular place or at a particular time educators must plan in advance to ensure that all material is available and assessed in a timely manner. Educators must also make sure the person getting credit for the class is, indeed, the person who does the work in the class. But perhaps the most important concern is whether or not online courses offer learning opportunities that are comparable in quality to traditional, face-to-face courses. Such assessment is notoriously difficult to conduct.

While many educators have offered various opinions of the efficacy of online classes, there is, as of yet, no definitive ruling on the value of online learning relative to face-to-face learning. Numerous factors impede progress in our understanding. First, there is no concrete definition of what it means for a class to be an "online" class. For some, it means that some ancillary content such as lecture notes or practice quizzes reside in an electronic format easily accessible to students while the classroom itself remains in the traditional format. For some, it means that all content – lecture videos, PowerPoint slides, class notes, quizzes, chat rooms – exist exclusively in electronic format. Various mixes of the approaches are legion. Of most interest to many researchers are the forms of online teaching that can be thought of as complete substitutes for the face-to-face format.[i] Second, it is very difficult to devise an

experiment that isolates the effect of having a class online relative to a traditional face-to-face class. The research community in general frowns upon (with good reason) using students as test subjects without imposing strict conditions to protect the welfare of the students involved. Therefore, strict laboratory experiments are pretty much out of the question. Nevertheless, to test the efficacy of the online delivery format one would want to avoid asking students to volunteer to take the online class as opposed to the face-to-face class. Given the choice, most students would gravitate toward the class format in which they believe they are most likely to excel. This self-selection problem will bias any comparison between the two venues. In fact, an extensive literature search conducted by the U.S. Department of Education in an effort to summarize the research concerning the efficacy of online delivery of course content found no experimental studies prior to 2006 with sufficient design and data gathering techniques to qualify as a truly random draw study (Means et al., 2010). In this study we describe a protocol for constructing a random assignment experiment that we hope will be a model for others to replicate.

While student self-selection is a hurdle, so is instructor self-selection. Just as students would normally gravitate toward the course venue in which they expect to do the best, instructors tend to gravitate toward their relative strengths given the opportunity. It is difficult to compare student outcomes when you cannot control for instructor input. We address one facet of this problem in that the same professor teaches both sections of the course. We do not address the problem completely, however, because with only one professor we cannot tell how much of the observed effects of differing delivery methods are due to characteristics unique to him. Therefore, we hope to encourage others to replicate our study in which the same professor teaches students in both venues. Individual instructor characteristics will become less of a confounding factor through the accumulation of multiple replications of the study.

This study proceeds with a review of the existing work in the area and highlight some of the areas in which this work offers some advances in our understanding of the efficacy of online education. The following section presents a workable protocol to determine the role online education can play in higher education. The second section addresses the selection issue present in most previous studies and describe our approach to randomizing the assignment into control and treatment groups.[ii] The third section describes the course set up and the data gathering process. The next section provides a summary and analysis of the data. A discussion of the results and conclusions wrap up the study. The results demonstrate some evidence that the mechanism researchers use to compare the performance of online students to traditional face-to-face students may be an important driver of their results. Furthermore, there is reason to doubt that the two pedagogical approaches are interchangeable.

Literature Review

There is a media/technology literature that predates the development of online delivery. Many of these early studies found a "no effect" result when comparing the learning outcomes of different media. According to Clark (1983), one should not expect the choice of media to have any effect on learning outcomes. He argues that those studies that do find different outcomes are actually picking up the effect created when an instructor switches to a new medium and must re-evaluate how the course material is presented within the new medium (Clark, 1983)[iii]. In other words, there is a change in the method of instruction as well as a change in the medium which is confounding the results. Kozma (1994) reframes the debate by acknowledging that each medium has specific attributes which are conducive to certain types of mental processes and social interactions. Clark's position is that the medium does not

matter, other things being equal. The Kozma position is that it is not useful to hold other things equal. The Kozma framework is more relevant to the current debate since it is commonly recognized that online delivery changes the nature of the relationship between the student and fellow students and between the student and the instructor. For example, the initial media literature focused on the difference between video and text. Such a comparison is not that useful now since both online and face-to-face delivery would employ a combination of video and text media. In fact, contrary to what many students may expect, online delivery is often more text based than face-to-face delivery. Therefore, we cannot assume that previous conclusions hold in the current environment.

Initially, the online environment was used as a tool to augment the traditional delivery of classroom material. Consequently, there are many studies that demonstrate that utilizing online material in addition to lecturing has positive benefits. In one such study, Coates and Humphries (2001) show that having online material available is useful but stress the importance of students actively engaging in the material for it to have its full effect. Passive interaction (such as reading other students' posts) had little impact on student performance. Van der Merwe (2011) is another good example of studies that show the positive effect of supplementing a traditional classroom experience with an online component. While there is fairly strong consensus that additional resources will generally improve outcomes (Means et al., 2010), such results are neither surprising (although, diminishing marginal returns ought to kick in at some point) nor particularly informative concerning the comparative efficacy of online delivery. More inputs, provided that their marginal productivity is positive, should generally lead to an increased measure of output – but at an increased cost of production.

Early Comparisons: No Random Selection

To evaluate the effectiveness of the online course format compared to the face-to-face course format one needs an experimental design that provides a falsifiable hypothesis and the means to test the hypothesis. When examining the outcomes of students who chose which class to attend, it is difficult to separate the impact of course design from the effect of the student's course preferences. The early economics literature that examines the evolution of the online course format attempts to control for selection issues by modelling the choice of class venue using observable student characteristics. Outside of experimental and behavioural economics, economists rarely use data that is generated from the controlled environment of an experiment. Consequently, Heckman (1979) corrections for selection bias and two-stage least squares approaches are commonly accepted as reasonable second best ways to address the problem when random selection is not available (See for example Coates et al., 2001).

Those studies that directly compare online delivery to face-to-face delivery provide conflicting evidence. For example, Coates et al. (2001) show that the online format generally results in lower test scores. They find, however, that for those students who choose the online format, it probably resulted in higher grades than they would have achieved had they taken the same course in the traditional classroom. Their analysis points to a problem that bedevils much of the early research in the area: Most studies rely on data gathered from different sections of a class where students choose which section they would rather attend. Coates et al. show that the population of students that volunteers to take an online class is systematically different than those who choose to register for the face-to-face class.

In fact, one would expect the rational student to gravitate toward the venue in which she anticipates being more successful.[iv] Cao and Sakchutchawan (2011) for example,

find evidence both in their review of the existing literature and in their own study that female students, older students, working students, part-time students and students with family obligations tend to gravitate toward online courses at higher rates than their counterparts. This provides some evidence that, given the choice, different types of students will select modes of pedagogical delivery they anticipate will best suit their needs. And to the extent that these characteristics are linked to expected success in an academic discipline, one would expect selection bias to be a confounding issue.

The evidence in Coates et al. (2001) also suggests that students do, in fact, self-select into the different types of class, but not always to their advantage. Such self-selection, Coates and Humphries (2001) conclude, may put some students, particularly younger students, at a disadvantage compared more mature students. They find younger students systematically underperform relative to their older counterparts in the online setting. This result is disconcerting, they note, because the rapid increase in the use of online classes in principles and introductory classes primarily targets freshmen and sophomores.

Two other economic papers of note grapple with the issue of comparing the online to the face-to-face classroom in the absence of an experimental setting. Navarro and Shoemaker (2000) examine two different classrooms, an MBA level class and an undergraduate Principles of Macroeconomics class. Their intent is to examine course design, research design and the research results of their comparison of the two educational approaches. They note that they are constrained by their institution from randomly assigning students to the two different types of class. Instead they examine the different characteristics of the groups of students who self-select into the online format as compared to those who self-select into the face-to-face format. They find that for their MBA students the face-to-face students outperform the online students in all measures of performance (but most differences are not statistically different). But, for the undergraduate students, the one measure of comparison used (a set of 15-question short essays) showed the online students outperforming the face-to-face students. The difference was statistically significant at the 99% level.

The second paper of note is Brown and Liedholm (2002) which examines three different approaches to economic education: the traditional live approach, a hybrid approach where lectures are augmented with online material, and a strictly online approach. They focus on the impact student characteristics have on measured performance in the different modes of delivery. But, they cannot control for student selection. They find that students in the live classroom score significantly better than those in the online class and slightly better than those in the hybrid class.[v]

It is also worth pointing out that Brown and Liedholm (2002) include an analysis of two different levels of questions. Basic knowledge questions – definitions and recollection questions – prove no more difficult for the online students than for the face-to-face students. More advanced questions – application and analysis questions – were measurably more difficult for the online students than the face-to-face students. Taken together with the Coates et al. findings and the concerns of Coates and Humphreys (2001), the results of Brown and Liedholm (2002) suggest a tension between the students who are most likely to succeed in an online class and the material best suited for an online class. Introductory material seems best suited for online classes, while more mature students seem better suited for online classes.

Studies with Random Selection

Missing from the discussion are studies that approach the problem from a true experimental angle. As many of the existing studies note, it is difficult to create an environment in which students are randomly assigned into a face-to-face classroom

and an online environment. For all of the careful attempts to control for selection bias, very few studies eliminate it through their experimental design. We are aware of three: One exception is Figlio, Rush and Yin (2013). They design an experiment where they randomly assign registrants for one large section of an introductory microeconomics course into an online section and a live section. To entice participation they promise students who participate in the experiment a half-letter grade boost in their final course grade. They monitor student attendance to prevent students who are in the online section from attending live lectures. And, they modify the computer access of students registered for the live section to impede them from viewing the material intended for the online students. They compare the students who volunteer for the experiment to those who chose not to participate and find that they are statistically similar in all but three measured characteristics. The volunteers had higher GPAs but lower SAT scores than the nonparticipants. Also, the volunteers were about ten percentage points less likely to have a mother who graduated from college.

The online class consisted of video tapes of the live lectures in addition to supplemental material common to both section. One concern of the authors is that although they restrict the students in the live section from accessing the online lectures from their student accounts, they cannot monitor the students' use of other computers. In their overall assessment of the experiment Figlio et al. find no statistical difference between the performances of students in the two sections. But, when they look at some of the sub-groups of students they find some evidence that students of Hispanic descent perform better in the face-to-face classroom. They also find males and low-achievers do better in the face-to-face section.

A second exception is Alpert, Couch and Harmon (2015). They randomly assigned students into three separate sections: face-to-face, blended and online. Their careful and thorough analysis found that online students under-perform relative to students in the other two formats. Their measure of student performance was the score on a cumulative final exam, and they conducted the experiment over four consecutive semesters. Students in the online sections scored five to ten points below those in the other sections. They also found that disadvantaged students in both the blended and online sections do worse relative to those in the face-to-face section.

A third exception is the study of Joyce, Crockett, Jaeger, Altindag and O'Connell (2014). They also manage a random selection of students into two class formats. They examine, however, the impact of more classroom time rather than a strict comparison of an online format to a traditional face-to-face format. Their two groups of students have access to the same online material but one group has two 75-minute face-to-face classes per week while the other group has only one 75-minute face-to-face class per week. Although the students with more face-to-face time with the professors do better on the measures of students' success (an accumulation of test scores throughout the semester), they conclude that the difference is small enough to justify the substitution of online material for face-to-face classroom time.

Randomized Enrolment Procedure

This study develops an approach that addresses some of the critical shortcomings of many of the existing studies. First, like Figlio et al. (2013) we randomly assign students to two sections taught simultaneously by the same professor. Unlike Figlio et al. the online section is not just video tapes of lectures. Rather, the professor designed the section to be specifically online. This includes the use of online mini lectures to provide material that would normally be presented in face-to-face lectures and discussion-based interaction that strives to create a sense of community among

the students and instructor. The idea behind this type of online course design is that optimal learning can only "take place when a student is actively involved within a social context (Bender, 2012, page 23)." A second difference is that we pre-test and post-test students using questions from the macroeconomics Test of Understanding in College Economics (TUCE). This second modification allows us to examine the different results between testing of instructor-specific material and standardized material.

Administratively, the main challenge was to randomly assign students into the two sections. Since this required an enrolment process quite different from the norm, the cooperation and collaboration of the university Registrar was essential. Since our study involved student participation and student data, we also needed approval from our Institutional Review Board (IRB) before we could begin. This required submitting a form which described the nature and methodology of the study, as well as how the study would address potential ethical issues (See Exhibit 2.). We did not seek exempt status since our study included student data beyond their in-class performance. After receiving approval to conduct the experiment from our campus IRB, we sent an e-mail to all business majors who had not yet taken Principles of Macroeconomics, inviting them to participate in random enrolment into one of two sections, one of which would be online.[vi] The only incentive offered for participation was a guaranteed spot in the class. Since this is a required course for all business majors, course sections typically fill up quickly. Waitlists persist and often do not clear. Excess demand aside, participation was slightly lower than we originally anticipated. The small class size of both sections may indicate a different source of selection bias: those students who agree to participate in the experiment may not be representative of the student body. Ideally, we would have offered more of an incentive to participate in the experiment.

The invitational e-mail included a link to a secure website along with a password where students could sign up for random enrolment. The e-mail and website made it clear that by signing up they were giving us permission to make anonymous use of their confidential administrative data. The e-mail also explained that if students were unhappy with their section assignment they could drop the class, but they would not be allowed to switch to the other section.[vii] Any such switches would obviously introduce selection bias into the experiment. Throughout the semester only two students dropped the class (one from each section). Three additional students, however, appeared to have effectively dropped the course since they stopped taking tests and participating in class.[viii]

The e-mail and website also informed the students of the days and times that the face-to-face class would meet so they could keep that time open when registering for other classes. Everyone participating in the study was required to meet on the first day of class to be randomly assigned to each section, to take a pre-test and to receive instructions for the rest of the semester. To assign sections the instructor printed each student's name on a strip of paper and placed it inside a basket. The instructor then went around the room having each student draw names from the basket. Students were alternatively assigned to the online and face-to-face section based on the order in which names were drawn. One down-side of waiting until the beginning of the fall semester for section assignment was that some students were anxious to know their assignment and e-mailed the instructor throughout the summer asking about the timing of random assignment. It is understandable that some, if not many, students disliked the uncertainty of not knowing their section assignment. Another possible inconvenience was that students who ended up in the online section would have been able to register for a class during the face-to-face class time if they had known their section assignment sooner. On the other hand, the approach limited the

ability of students to drop the course with enough time to enrol in an alternative course if they did not like the outcome of the drawing.

Course Structure

Aplia by Cengage was the learning platform used for both sections for exams, homework, course materials and class announcements. The exam grades comprised 65% of the course grade for both sections. The homework grade was 20% of the overall grade for both sections. The exams and homework assignments were the same and were administered through Aplia for both sections. The remaining 15% varied across sections. Students in the online section were required to respond to weekly discussion questions. The instructor would occasionally post a response to guide and focus the discussion. The instructor would also provide more extensive feedback when the discussion was closed for each question. The main purpose of discussion questions in online classes is to replace the intellectual engagement and sense of community that comes with in-class, face-to-face interaction. (See Hammond (2005) for a review of the purported benefits of asynchronous online discussion.) Students in the face-to-face section were assigned three short papers in place of online discussion.

The mini lectures were available to the online section only. Most of these mini lecture were short Word documents written by the instructor, usually one and a half to two and a half pages in length. A few were power point slides with accompanying audio recordings by the instructor. These mini lectures, along with the instructor's commentary and feedback to the discussion question responses, were meant to substitute for face-to-face lectures. The reading list for the online class contained numerous links to short videos that helped illustrate important concepts. The instructor also showed most of these videos to the face-to-face class during lectures.

Method

The Registrar's Office provided student administrative data. Table 1 presents data for students in the class as a whole, those in the face-to-face section and those in the online section. The characteristics of the two sections closely mirror each other (granted, the class itself is fairly homogenous to begin with) which demonstrates one of the important aspects of the random selection process – many of the differences among students that characterize past studies is not present. One observation worth noting, however, is that the students in the online section seem to have slightly better measures of human capital prior to the class starting (math SAT equivalent scores, high school GPA and institutional GPA) than the face-to-face students. The differences are not big, but if they were to introduce any bias into the study it would probably be in the direction favouring the online section. Table 2 summarizes the stated majors of the students in each section. All but a few students had declared for various business majors.

Thirty-seven students agreed to participate in the study. One student from each section eventually withdrew for a total thirty-five students taking the course. However, three students stopped participating in the class without officially withdrawing from the course. Consequently, thirty-two students enrolled and completed the course. Seventeen were enrolled in the face-to-face section and fifteen in the online section.

Table 1:

Descriptive statistics

	Total n=32	s.d.	FtF n=17	s.d.	Online n=15	s.d.
SAT Math	546.88	64.83	541.18	68.82	553.33	61.72
GPA*	3.26	0.41	3.21	0.45	3.32	0.36
% male	0.47	0.51	0.53	0.51	0.40	0.51
White	0.94	0.25	0.88	0.33	1.00	0.00
Non.white	0.06	0.25	0.12	0.33	0.00	0.00
Age (birth year)	1993.37	0.88	1993.28	1.02	1993.47	0.72
Inst. GPA	3.19	0.46	3.18	0.46	3.20	0.48
Exam Average	75.6	10.45	78.75	8.39	72.04	11.65
Total Change	10.59	4.36	10.58	3.79	10.6	5.07
TUCE Change	5.97	3.59	5.41	3.47	6.6	3.74
Instructor Change	4.63	1.91	5.18	1.91	4	1.77
* one observation missing for transfer student who is in the face-to-face class						

The study presents two measures of student performance. The first measure is improvement on the pre-test questions. The pre-test administered on the first day of class consisted of thirty multiple-choice questions. Twenty questions were drawn from the third edition of the TUCE exam, and the remaining ten were drawn from the instructor's old exams. Twenty-six of these questions reappeared on one of the three exams given throughout the semester. Four questions did not appear again because the corresponding material was not covered in the course. In hindsight, we would have liked the instructor to have used these questions again to serve as a benchmark and to examine the hypothesis that students benefit just from having seen a question before. The second measure of student performance is the overall percentage grade on the three course exams. The face-to-face section took the same online exams as the online class to eliminate differences in performance due to differences in exam format or administration. As previously stated, exam grades accounted for 65% of the overall course grade for both sections. Table 1 also presents the means and standard deviations for each measure of performance, including disaggregated results for instructor and TUCE questions separately.

An obvious question is if the differences in the mean performance between the online and face-to-face sections are statistically significant. Conducting two-tailed,

homoscedastic hypothesis testing for the mean differences in all four measures of performance yields mixed results. The higher mean exam average for the face-to-face class is statistically significant at the 90% level of confidence (p = 0.06908), as is the higher mean improvement for the instructor TUCE questions (p = 0.08242). There is no evidence to reject equality between means in the other two measures of performance.

Table 2:

Student majors

Majors	Total	FtF	Online
Management	11	7	4
Accounting	6	3	3
Marketing	6	2	4
Economics	3	2	1
Psychology	1	0	1
Mgmt. Info. Systems	2	1	1
Business, Undecided	1	0	1
Environmental Science	1	1	0
Pre-Mass Communication	1	1	0

Although the sample size is rather small, we model both the change in student scores from the pre-test to the post-test and each student's course exam average as a function of the environment in which the student took the course, the student's SAT math score (or ACT equivalent), the student's overall GPA prior to taking the course, the student's gender and the overall credit hours the student had earned prior to taking the course. Using ordinary least squares analysis, there are results from four linear regressions based on estimating the following equation.

$$Performance_i = \beta_0 + \beta_1 Online_i + \beta_2 SATM_i + \beta_3 GPA_i + \beta_4 Male_i + \beta_5 Hours_i + \varepsilon_i$$

Results

The dependent variable for the first regression is the difference from the pre-test to the post-test for all of the questions of interest. The second is the difference from the pre-test to the post-test for the questions culled from the TUCE exam. The third is the difference from the pre-test to the post-test for the questions chosen by the instructor. Finally, the fourth dependent variable is the exam average for each student.

Table 3:

Dependent Variable = Total Post-test – Pre-test

Multiple R	0.692		
R Square	0.479		
Adjusted R Square	0.379		
Standard Error	3.437		
F	4.783		
Significance F	0.003		
Observations	32		
	Coefficients	*Standard Error*	*t Stat*
Intercept	14.996*	7.750	1.935
On-line	-0.270	1.244	-0.217
SAT Math	-0.005	0.010	-0.496
Overall GPA	1.538	1.569	0.980
Male	2.285*	1.258	1.816
Overall Hours Earned	-0.169***	0.039	-4.340
* 90% confidence level, ** 95% confidence level, *** 99% confidence level			

Table 4:

Dependent Variable = TUCE Post-test – Pre-test

Multiple R	0.617		
R Square	0.381		
Adjusted R Square	0.262		
Standard Error	3.082		
F	3.202		
Significance F	0.022		
Observations	32		
	Coefficients	*Standard Error*	*t Stat*
Intercept	5.352	6.948	0.770
On-line	0.961	1.115	0.862
SAT Math	-0.004	0.009	-0.404
Overall GPA	1.901	1.407	1.351
Male	1.768	1.128	1.567
Overall Hours Earned	-0.108***	0.0349	-3.087
* 90% confidence level, ** 95% confidence level, *** 99% confidence level			

Table 5:

Dependent Variable = Instructor Post-test – Pre-test

Multiple R	0.604		
R Square	0.364		
Adjusted R Square	0.242		
Standard Error	1.666		
F	2.980		
Significance F	0.029		
Observations	32		
	Coefficients	*Standard Error*	*t Stat*
Intercept	9.644**	3.756	2.568
On-line	-1.231*	0.603	-2.042
SAT Math	-0.001	0.005	-0.275
Overall GPA	-0.363	0.761	-0.477
Male	0.517	0.610	0.849
Overall Hours Earned	-0.061***	0.019	-3.245
* 90% confidence level, ** 95% confidence level, *** 99% confidence level			

Table 6:

Dependent Variable = Exam Average

	Coefficients	Standard Error	t Stat
Multiple R	0.702		
R Square	0.493		
Adjusted R Square	0.395		
Standard Error	6.814		
F	5.056		
Significance F	0.002		
Observations	32.000		
	Coefficients	*Standard Error*	*t Stat*
Intercept	45.1***	15.363	2.936
On-line	-4.164*	2.466	-1.689
SAT Math	0.001	0.020	0.064
Overall GPA	12.164***	3.111	3.910
Male	2.942	2.495	1.179
Overall Hours Earned	-0.136*	0.077	-1.759
* 90% confidence level, ** 95% confidence level, *** 99% confidence level			

A consistent result is that hours earned is significantly different from zero at the 1 percent level for the first three regressions and the 10% level for the final regression. Interestingly, the sign is negative. Because the course is a principles level course and should be taken early in a student's college career, a higher value for hours earned may be an indication of a student being less diligent in progressing appropriately. The coefficient on Male in the first regression indicates males have a roughly two-point advantage in the comparison of pre-test scores to post-test scores. This advantage is no longer statistically significant when the comparison questions are disaggregated into the TUCE questions and instructor questions. The coefficient on Male is also not a significant factor in determining the exam average. The main result is that in both the model of the comparison of pre-test and post-test performance on the instructor questions and the model of overall exam average, the coefficient for Online is statistically significant and negative. In the case of the comparison of pre-test and post-test scores on the instructor questions, taking the online course would be expected to decrease the change in score by more than a point. This represents roughly 25 percent of the average change in score. And for the overall grade, an expected decrease in exam average of 4.164 percentage points represents almost half

a letter grade. Keeping in mind the small sample size makes the estimates less reliable, we find these results interesting. These mixed results point to the importance of making clear how one intends to measure the effectiveness of different teaching modes.

Discussion

We were careful to note situations in which we thought the experiment could be improved. Following are situations in which we think further studies could shed light on whether or not problems exist in the design.

There were many students in the face-to-face section who knew someone in the online section, so the possibility of cross-contamination does exist. We could not gauge the extent to which face-to-face students had access to the online material (the sample was small enough to prevent online students from attending the live lectures).

The instructor noticed that the face-to-face section had unusually good attendance relative to a typical face-to-face section of macro principles. This may have been a random occurrence or it could have been an indication of the Hawthorn effect in that students, knowing they were in a study, altered their typical responses. It could also be due to the class being smaller than average for a Principles of Macroeconomics class. Perhaps students in smaller class feel less anonymous, and therefore are more likely to attend class.

Because the students had to agree to partake in the experiment in advance, there may be a selection bias arising from the possibility that the willingness to participate in randomized enrolment could be correlated with other student characteristics associated with learning and academic performance. That is, those students willing to participate may be systematically different from the general student population in some meaningful way. If this is the case it would reduce the ability to generalize from the observed results.

We did not have an exit survey ready for when students asked to withdraw from the class. In both cases, the survey was sent after the withdrawal form was signed. Although both students verbally agreed to complete the survey, neither one did. It would be beneficial to know why students withdrew from the experiment. Unfortunately, there was no way in our institution to force the students to fill out the exit survey.

Participation was lower than we would have liked. Typically, a section of Principles of Macroeconomics at our institution will enrol 45 to 65 students. We believe that the low participation numbers were due to the small incentive (although our students have average-time-to-completion rates that are now above 5 years, perhaps they do not consider early completion of core courses to be a sufficient benefit). A more generous enticement may be required, or a more thorough advertisement of the experiment might be sufficient. We doubt, however, that the seemingly generous offer of one-half of a letter grade as offered by Figlio et al. (2013) for participation in their study would be authorized at our institution. As we pointed out earlier, those students who agreed to participate in the experiment may have different characteristics than students in a typical section of the class. This may create a selection bias in the results.

One of the errors we made (as noted above) was we neglected to post-test some of the TUCE questions that pertained to topics that were not covered during the semester. It would have been useful to post-test those questions (and not count the results toward the final grade) in order to determine what level of learning was

occurring simply because students had seen certain questions in the pre-test. In larger studies one could then better control for the background level of human capital development.

Finally, we did not gather information from students on their prior experience with online classes. It would be useful to know how many online or hybrid courses each student previously had and the structure of the courses. Similar information for instructors would be relevant in a large study with multiple instructors.

Conclusion

Students were randomly assigned to either an online or face-to-face section of the same class to minimize sample selection bias. There were two broad measures of student performance: the exam average for the entire course and the improvement on a post-test relative to a pre-test. The pre- and post-test had both standardized TUCE questions and instructor questions. The face-to-face class performed statistically, significantly better than the online class in terms of the exam average and improvement in post-test instructor questions. There was no statistical evidence for a difference in improvement in total post-test questions or improvement in post-test TUCE questions. These mixed results remain when using ordinary least squares analysis to control for the student's SAT math score (or ACT equivalent), the student's overall GPA prior to taking the course, the student's gender and the student's overall credit hours prior to taking the course

The sample size does not allow for anything more than a suggestion that the two approaches in question may led to different results depending on how one measures student performance. We hope to motivate others to replicate the experiment in different disciplines and institutions in an effort to build a systematic assessment of the relative merits of online learning as compared to the traditional face-to-face classroom. There are a number of interesting anecdotes that arise from this study that are worth more meticulous examination. For one, the difference in relative performance on the standardized questions and the instructor specific questions suggest two different issues: 1) Are some levels of learning better facilitated by online classes than others, as suggested by Brown and Liedholm (2002)? And 2) are there nuanced learning objectives in classrooms that can only be transmitted in a face-to-face environment? By incorporating both questions from the TUCE and instructor-specific questions in a pre and post-test format, we believe these issues can be reasonably addressed. A second line of inquiry is raised by the fact that our data is relatively homogeneous to a number of student characteristics. We cannot, for example, examine issues of race or ethnicity due to the lack of variation in the data. Yet, studies such as Figlio et al. (2013) find that the format of the class has a much different impact on Hispanic students than it does on other groups of students. An examination of data from other environments would allow researchers to revisit the findings of Figlio et al. (2013).

Hopefully, other researchers will find this an instructive study, and they will pick up the challenge to definitively test the hypothesis that the online environment is a reasonable substitute to the face-to-face environment. While this study examines the impact of the delivery method on students in a Principles of Macroeconomic class, instructors from any discipline with a standardized testing instrument could provide additional insight.

Acknowledgements

References

Allen, I. Elaine & Seaman, J. (2006) *Making the Grade: Online Education in the United States*. Massachusetts: Sloan Consortium.

Alpert, W. T., Couch, K. A. Harmon, O. R. (2015). *A Randomized Assessment of Online Learning*, Unpublished Manuscript, Department of Economics, University of Connecticut, Storrs, CT.

Bender, T. (2012). *Discussion-based Online Teaching to Enhance Student Learning*, 2nd edition. Sterling, Virginia: Stylus.

Brown, B. W. & Liedholm, C. E. (2002). Can Web Courses Replace the Classroom in Principles of Microeconomics? *American Economic Review, (Papers and Proceedings)*, 92(2), 444-448.

Calderon, V. J. & Sidhu, P. (2014). Business leaders say knowledge trumps college pedigree. Gallup 2/25/14 (accessed 9/30/14: http://www.gallup.com/poll/167546/business-leaders-say-knowledge-trumps-college-pedigree.aspx?version=print).

Cao, Y. & Sakchutchawan, S. (2011). Online vs. Traditional MBA: An Empirical Study of Students' Characteristics, Course Satisfaction, and Overall Success, *Journal of Human Resource and Adult Learning,* 7(2), 1-12.

Clark, R. E. (1983). Reconsidering Research on Learning from Media, *Review of Educational Research*, 53(4), 445-459.

Coates, D. & Humphreys, B. R. (2001). Evaluation of Computer-assisted Instruction in Principles of Economics, *Education Technology & Society*, 4(2), 133-144.

Coates, D., Humphreys, B. R., Kane, J., Vachris, M., Agarwal, R., & Day, E. (2001). 'No significant distance between face to face and online instruction: Evidence from principles of economics. Working paper.

Figlio, D. N., Rush, M. & Yin, L. (2013). Is it Live or is it Internet? Experimental Estimates of the Effects of Online Instruction on Student Learning, *Journal of Labour Economics* 31(4), 763-84.

Frederickson, N., Reed, P., & Clifford, V. (2005). Evaluating Web-supported Learning versus Lecture-based Teaching: Quantitative and Qualitative Perspectives, *Higher Education* 50, 645-64.

Gratton-Lavoi, C. & Stanley, D. (2009). Teaching and Learning Principles of Microeconomics Online: An Empirical Study, *Journal of Economic Education* 40(1), 3-26.

Hammond, M. (2005). A Review of Recent Papers on Online Discussion in Teaching and Learning in Higher Education, *Journal of Asynchronous Learning Networks*, 9(3), 9-23.

Harmon, O. R. & Lambrinos, J. (2007). *Student Performance in Traditional vs. Online Format: Evidence from Introductory Economics Classes*, University of Connecticut, Department of Economics Working Paper Series 2007-03R.

Heckman, J. (1979). Sample Selection Bias as a Specification Error, *Econometrica,* 47(1), 153–61.

Hoxby, Caroline M. (2014). *The Economics of Online Postsecondary Education: MOOCs, Nonselective Education and Highly Selective Education*, NBER Working Paper No. 19816.

Hoxby, Caroline M. (2017). *The Returns to Online Postsecondary Education*, NBER Working Paper No. 23193.

Joyce, T. J., Crockett, S. Jaeger, D. A., Altindag, O. & O'Connell, S. D. (2014). Does Classroom Time Matter? A Randomized Field Experiment of Hybrid and Traditional lecture formats in economics, NBER Working Paper No. 20006.

Kozma, R. (1994). Will Media Influence Learning? Reframing the Debate, *Educational Technology Research and Development*, 42(2), 7-19.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K (2010). *Evaluation of Evidence-based Practices in Online Learning: A Meta-analysis and Review of Online Learning Studies*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development.

Navarro, P. & Shoemaker, J. (2000). Policy Issues in the Teaching of Economics in Cyberspace: Research Design, Course Design, and Research Results. *Contemporary Economic Policy*, 18(3), 359-366.

Test of Understanding in College Economics – 3rd edition (1991) Joint Council on Economic Education New York, NY.

van der Merwe, A. (2011). Can Online Learning Boost Academic Performance? A Microeconomics Study, *The International Business & Economics Research Journal,* 10(8), 45-55.

World Economic Forum (2016). *New Vision for Education: Fostering Social and Emotional Learning through Technology*, (accessed 10/18/2017: https://www.weforum.org/reports/).

Zhang, D. (2005). Interactive Multimedia-based e-Learning: A Study of Effectiveness, *American Journal of Distance Education*, 19(3), 149–62.

Zhang, D., Zhou, L., Briggs, R. O. & Nunamaker Jr., J. F. (2006). Instructional Video in e-Learning: Assessing the Impact of Interactive Video on Learning Effectiveness, *Information and Management*, 2006, 43(1), 15–27.

Appendix

Exhibit 1: Email to students

E-mail for Students: "On-line versus Face-to-face Instruction: A Pilot Study"

Subject: Guaranteed and randomized enrolment for Econ 2105 – Principles of Macroeconomics

Dear Student,

You are invited to participate in random section enrolment for Econ 2105 in the fall semester of 2013. This is part of a University research study examining the relative effectiveness of on-line instruction. The benefit for you is a guaranteed seat in the class. If you agree, you will be randomly enrolled in one of two sections: a traditional face-to-face class that meets twice a week or an on-line class. If you agree, there is a 50 percent chance that you will be enrolled in the on-line section. This will be the only on-line section of Econ 2105 – Principles of Macroeconomics – offered in the fall and spring semesters. Both sections will be taught by the same instructor, and will be capped at 45 students. The face-to-face class will be on Tuesday and Thursday from 11:00 am until 12:15 pm.

You are being invited to participate in this study because you are a rising sophomore and you have indicated that you plan to be a business major. Econ 2105 is a prerequisite for many business courses and is required in Area F for all business majors.

The on-line class will meet face-to-face only once at the beginning of the semester. If you are unhappy with your assigned section, you may drop the class. However, you will not be able to switch to the other section involved in the study.

"Participation" entails allowing us to use the anonymous data generated by the study and a possible exit survey. If you later choose not to participate in the study, you can still remain in your assigned section of the class. Refusal to participate will not affect your course grade in any way.

If you are interested in participating, please go to GCSUECON2105 to indicate your willingness to participate in the study and to sign-up for randomized section assignment. Your password to access the website is *GCSUmacroprinciples*. Once in the website, you will be asked to enter your GCID. Afterward, you will receive a follow-up e-mail to confirm your acceptance and enrolment.

I hope you will choose to participate in this important study as the results will help us better understand the relative merits of offering on-line courses relative to traditional face-to-face classroom setting.

Thank you,

Exhibit 2: IRB Form

GEORGIA
COLLEGE
REQUEST TO CONDUCT STUDIES WITH HUMAN
SUBJECTS

This sample form may be used to prepare to submit a new protocol to the Georgia College IRB. All submissions are completed online (irb.gcsu.edu) as of January 2010. DO NOT SUBMIT THIS FORM.

Annual renewals for a previously approved protocol are submitted using the form included in your approval email.

1. Protocol Title:

2. Date Research will begin: Date of expected completion:

3. Principal Investigator & Administrative Contact (individual responsible for completing paperwork): Name: Email:
    Phone: Campus Box or Mailing
    Address: Department/section:

4. Supervising Faculty (for student conducted research): Name: Email:
    Phone: Campus Box or Mailing
    Address: Department:

5. Associated Personnel
    Name: Email:
    Phone: Campus Box or Mailing
    Address: Department:

    Name: Email:
    Phone: Campus Box or Mailing
    Address: Department:

6. Sponsor (funding source):

7. Location of Research (place an X in the brackets of all that apply):
    [ ] Georgia College & State University   [ ] River Edge Behavioral Health Center
    [ ] Oconee Regional medical Center   [ ] Medical Center of Central Georgia
    [ ] Northside Hospital   [ ] HCA Coliseum Hospital
    [ ] Central State Hospital   [ ] The Methodist Home
    [ ] Georgia School for the Blind   [ ] Charter Hospital
    [ ] Other:

8. Tissue only: [ ] Yes   [ ] No
(please note: discarded tissue – no identifying link to subject and/or no possibility of need of cell line waiver.)

9. Subject sex: [ ] Male   [ ] Female   [ ] Both

10. Subjects' age: [ ] Infant/Toddler (0-2 years)   [ ] Child (3-12 years)   [ ] Adolescent (13-19 years)
                    [ ] Adult (20+ years)   [ ] Geriatric (65+ years)

11. For initial submission, please note the expected number of subjects to be enrolled in the investigation:

a. Total # of subjects included on-campus. b. Total # of subjects included off-campus. c. Total # of subjects included at all centers.

12. What type of study is proposed?

[ ] Survey   [ ] Retrospective (case-control)   [ ] Community Intervention
[ ] Pilot Study   [ ] Cohort (longitudinal) study   [ ] Laboratory Experiment
[ ] Clinical trial   [ ] Multi-center investigating   [ ] Program/policy study
[ ] Cross-sectional   [ ] Compassionate use
[ ] Other: _____

13. Keywords (used to describe the research in this protocol): _____

14. I certify that this protocol conforms to the OSHA/HHS guidelines for HIV/HBV occupational safety.

_____
Signature of Principal Investigator   Date

15. Description of the Research (provide sufficient detail):
    a. Statement of the problem:

    b. Data collection methods:

    c. Instruments to be used: (attach new or nonstandard ones, including researcher generated surveys.)

    d. Method of recruitment of participants: (send in any advertisements)

    e. Incentives, follow ups, compensation to be used:

    f. Detail stress, psychological, social, legal, or physical harm that might occur to participants. How are these held to the absolute minimum? What remediation is offered?

    g. Benefits of the research: University policy requires that any risk associated with participation be outweighed by potential benefits to participants and to humankind in general.
        a. Identify any benefits to participants resulting from this research:

        b. Identify any benefits to humankind in general resulting from this research:

    h. Consent Forms: How will legally effective informed consent be obtained from all participants (or their parent(s) or guardian(s))? Include form(s) to be used. (A model for appropriate informed

consent is available at the IRB website.) If deception is necessary, please justify, and describe and submit debriefing procedures.

i. Minors and others: If minors or other vulnerable participants are involved, please outline procedures to be used in obtaining their agreement (assent) to participate, in addition to the consent of the parent(s) or guardian(s). An assent agreement, similar to informed consent, must be obtained from children and adolescents ages 12-18 years.

j. Future Risk: How are all participants protected from the potentially harmful future use of the data collected in this research? Describe measures planned to ensure anonymity or confidentiality. If audio or videotapes are used, when will they be erased?

k. Illegal Activities: Do the data to be collected relate to illegal activities? If so, please explain.

[i] Our institution has different definitions for various mixes of face-to-face and cyber student experiences. A course is considered to be fully online if 95 to 99% of the student experience is online.

[ii] Technically speaking, the experiment is not truly random in that student subjects agreed in advance to participate in the experiment as described below.

[iii] He also mentions a bias in favor of new technology and the use of different instructors across media. Neither of these is a factor in our study.

[iv] For some, if not many, students, this will mean selecting the format with the lowest cost grade. Another relevant factor is that some students and instructors believe it is easier to cheat in an online class.

[v] Although, this comparison should be taken with a grain of salt as the professor in charge of the class was different which introduces a selection bias on the part of the professor.

[vi] The completed and approved IRB form is available to interested parties.

[vii] Students are limited in the number of times that they can drop classes after the add/drop period. This constraint is typically not yet binding for students still taking Principles level classes.

[viii] The University has a liberal grade forgiveness policy in which a student can erase a poor grade by retaking the course.