Economics of Education Review 46 (2015) 64–77

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Does classroom time matter?

Ted Joyce [a,b,c,*], Sean Crockett [a,c], David A. Jaeger [b,c,d], Onur Altindag [c], Stephen D. O'Connell [c]

[a] Baruch College, Department of Economics and Finance, 55 Lexington Avenue, New York, NY 10010, United States
[b] NBER, United States
[c] CUNY Graduate Center, Ph.D. Program in Economics, 365 Fifth Avenue, New York, NY 10016, United States
[d] University of Cologne, Germany

ARTICLE INFO

Article history:
Received 19 November 2014
Revised 16 February 2015
Accepted 23 February 2015
Available online 9 March 2015

JEL classification:
I20
I23

Keywords:
Classroom time
Microeconomics
Randomization
Student performance

ABSTRACT

Little experimental evidence exists on the causal impact of class time on academic performance when students have access to extensive course material online. We randomized 725 college students into traditional twice-per-week and compressed once-per-week lecture formats in introductory microeconomics. Students in the traditional format scored 3.2 out of 100 points higher (0.21 standard deviations) on the midterm than those in the compressed format but a statistically insignificant 1.6 points higher (0.11 standard deviations) on the final. There were no differences in non-cognitive outcomes. Students in the middle tercile of predicted test scores performed worst in the compressed format relative to those in the traditional format but there was little difference in test scores by format in the top tercile of predicted performance. While the compressed format offers clear savings in classroom space and professors' time, these savings come at some cost to student performance.

James A. Garfield, twentieth president of the United States and a graduate of Williams College, is reputed once to have said of renowned Williams educator Mark Hopkins: "the ideal college is Mark Hopkins on one end of a log and a student on the other" (Rudolf, 1956, p. vii). Garfield's epigram embodies the notion that the best learning takes place in a dialogue between student and professor, in which students take an active role in the learning process and professors can easily gauge a student's comprehension through verbal and non-verbal cues. This ideal remains at the core of American higher education despite the enormous changes in instructional technology that have occurred since the mid-19th century when Garfield was educated. In the mid 1950s, television was the first technology to capture the imagination of university administrators keen to reach a larger student population and, most importantly, hold the costs of instruction down (Eurich, 1958; Macmitchell, 1955). More recently, the Internet and various modes of online instruction have captured the imaginations of university administrators anxious to cut costs. Online learning in some form will surely be an increasingly important component of university education, even potentially improving on the kind of instruction Mark Hopkins might have offered to his students (Bowen, 2013).

To what extent does the opportunity to interact with a professor and other students matter in an environment rich in online materials? Recently, Figlio, Rush, and Yin (2013) compared students who took introductory economics online versus in a traditional lecture format at a major research university. Bowen et al. (2014) examined the performance of students in an introductory statistics class held on six

* Corresponding author at: Baruch College, Department of Economics and Finance, 55 Lexington Avenue, New York, NY 10010, United States. Tel.: +1 646 312 3541.
E-mail addresses: theodore.Joyce@baruch.cuny.edu (T. Joyce), Sean.Crockett@baruch.cuny.edu (S. Crockett), djaeger@gc.cuny.edu (D.A. Jaeger), raltindag@gc.cuny.edu (O. Altindag), soconnell@gc.cuny.edu (S.D. O'Connell).

http://dx.doi.org/10.1016/j.econedurev.2015.02.007
0272-7757/

public university campuses, contrasting the performance of students attending a traditional class with two weekly meetings with those whose class material was delivered online supplemented by one weekly class meeting. Both studies reported no overall difference in performance as measured by test grades between formats.[1] Participation rates in both studies were less than 25%, however, highlighting one difficulty of undertaking a classroom-based, semester-long randomized trial in a university setting.[2]

To gauge better the importance of classroom time in a typical "online rich" learning environment, we randomly assigned 725 students into "compressed" and traditional formats of introductory microeconomics at a large, urban, public university. We examine whether students who were offered class once a week for 75 minutes over a 14-week semester performed as well as students who were offered class twice per week, each for 75 minutes. Two experienced professors (the first two authors) taught four sections, one of each format. Students in the two formats had access to the same lecture slides, online material, and faculty-produced videos, which eliminated substitution bias as a source of attenuation since classroom time was the only difference between formats. Because research on student learning suggests that frequent assessments with immediate feedback improve performance (Pennebaker, Gosling, & Ferrell, 2013), we required students in both formats to take the online quizzes both before and after lectures using a sophisticated interactive web application (Aplia) to deliver and grade them.

We find that students in the traditional format performed 3.2 percentage points (p-value of 0.005) better on the midterm on a 100-point scale but a statistically insignificant 1.6 percentage points (p-value of 0.138) better on the final – differences of 0.21 and 0.11 standard deviations, respectively. Students in the lower tercile of predicted test scores performed worst in the compressed format on the midterm relative to those in the traditional format but students in the middle tercile performed worse in the compressed format overall. There was little difference by format in test scores in the top tercile of predicted performance. Students in both formats attended the same proportion of classes, and there were no differences in withdrawal rates. We also find no difference in hours logged into Aplia. Students in the compressed format watched 2.5 more videos than those in the traditional format relative to an overall mean of 8.5, while students whose professor was in the videos watched the videos 8.6 more videos than those whose professor was not in the videos.

Our results have meaningful pedagogical and administrative implications for undergraduate education. The fundamental difference in treatment between the traditional and compressed formats is the amount of time spent in the classroom, with students in the compressed sections having only half the amount of formal class time as those in the traditional sections. Differences in test scores by format were twice as large for the midterm relative to the final, suggesting that students in the compressed format adjusted to the demands of less class time. Nevertheless, reduced class time appears to diminish cognitive performance for most students in a large introductory economics class at a public university in which the vast majority of students commute.

1. The experiment

1.1. Setting

The study took place at Baruch College, part of the City University of New York and one of the most diverse campuses in the country. As of the 2013–2014 academic year, the Baruch student body claimed 163 nationalities and spoke 110 languages.[3] Baruch's Zicklin School of Business is the largest accredited collegiate school of business in the country with 12,000 undergraduates. Almost all students commute to campus and most attend full-time.

Principles of Microeconomics (ECO 1001) is a required course for all students applying to the business program at Zicklin. It also fulfills a social science requirement for non-business majors. Nearly 1000 students take ECO 1001 each fall. Four sections with seats for a total of 776 students were part of our study, which accounted for 95% of the daytime non-honors seats available for the course.[4] Students could register for class on Mondays and Wednesdays in the morning or Tuesdays and Thursdays in the late afternoon. Classes were listed as taught by the first two authors of the study. Both are full-time, tenured faculty members who have taught the class for the past six years and both have strong teaching evaluations.[5] Registration for the fall classes began in April of 2013 and continued through August. Students currently enrolled in Baruch could register in April and May while transfer students from community colleges or other four-year colleges could not begin registration until June.

[1] Figlio, Rush, and Yin (2013) did, however, find that Hispanic students and those with a grade point average below the median did less well in the online class.

[2] In addition to low participation rates, both studies encountered other difficulties. For example, Figlio, Rush, and Yin (2013) present results showing no mean differences in test scores between formats unadjusted for covariates but statistically significant differences of between 2 and 3 percentage points on a 100-point scale when adjusted. Students in the "live" format scored 3 percentage points higher on the final exam (p-value < 0.05) and 2.5 percentage points higher (p-value < 0.01) on the average of all three exams than students restricted to the video-taped lectures. See Table 3 in Figlio, Rush, and Yin (2013). The statistics experiment conducted by Bowen et al. (2014) encountered difficulty coordinating test and grading across campuses and faculty – not all campuses used a common set of questions on the final and faculty, aware they were part of an experiment, may have graded more leniently in order to reduce failure rates.

[3] Statistics about Baruch's student population are available from the authors upon request.

[4] Twenty-one seats went unfilled in the sections of the course in this study. Just over 100 students took ECO 1001 in the evening, most of whom were part-time students. Of the remaining students who were not part of our study, one section of 25 students was reserved for honors students only, and another daytime section of 40 students was taught by an adjunct faculty member.

[5] In student course evaluations for the fall semesters of 2010–2012, both professors averaged 4.4 on the 6 questions that assessed the quality of the course organization and delivery. Copies of the full teaching evaluations for each professor are in the Appendix. In addition, each professor has a rating of 4.3 based on a 1–5 scale of teaching ECO 1001. See http://www.ratemyprofessors.com (last accessed November 3, 2014).

1.2. The course

All sections of the class used N. Gregory Mankiw's *Principles of Microeconomics* (6th Edition) as the textbook (Mankiw, 2012), along with Cengage Learning's Aplia web application to administer and grade online quizzes. Each week students took a "pre-lecture quiz" due on Sundays and covering material to be taught in the upcoming week, and a "post-lecture quiz" due on Saturdays covering material that had been taught during the week. The pre-lecture quizzes were pass/fail (students who correctly answered at least half of the questions received full credit for the quiz) and were generally easier than the post-lecture quizzes; they were designed to ensure students came to lectures with some basic understanding of the material, without which the pace of the compressed lectures in particular would have been quite challenging for most students.

Lectures by professors formed the core of ECO 1001. During lectures, the professors presented microeconomic theory and examples using slides. The same slides were used in the compressed and traditional lectures by both professors, and were made available to all students for download, but they were covered more selectively and quickly in the compressed format, with less time to verbally annotate the slides, work through examples, and answer student questions. There was also less time in the compressed format to go over difficult problems from the Aplia quizzes and to review practice questions for exams.[6] In addition, one of the professors recorded videos for each chapter from the text, in which he annotated answers to 10 multiple choice questions. Each video was approximately 30 min long but was broken up into segments ranging from 5 to 10 min each, so that students could easily select only the videos for which they sought explanations. The videos were taped in a studio with no audience but the lighting and sound were professionally supervised.

In the once-per-week format the professors were forced to compress their lectures to fit the reduced class time, making the fundamental differences between the two formats a reduction in the amount of contact that students had with the professor (and classmates) and an increase in the pace of instruction during lectures. Although other studies have tried to limit access to online material, this struck us as infeasible and a potential confounding factor for the results. All online content was available to students in both formats of the class in order to isolate the impact of classroom time on student performance. We believe the contrast between the two formats in our study is likely to be closer to the "real world" implementation of such courses.[7] Moreover, because classroom space and time are far more costly to provide than online materials, our treatment captures the relevant margin on which university administrators are likely to prefer one format to another.

[6] While several practice exams and solutions were made available to all students online, the traditional lecture format presented more opportunities to visit the practice exams during class.

[7] Any attempt to limit access to online course material among students in the traditional format would likely have failed. Moreover, there is a plethora of free online material for introductory microeconomics currently available.

1.3. Recruitment and randomization

Recruitment began in May of 2013, shortly after the beginning of registration for the Fall 2013 semester. Students who had registered for one of the four class sections were sent an e-mail inviting them to participate in the study with a link to the electronic consent form. The CUNY Institutional Review Board, in approving our application, allowed us to offer an incentive of five extra-credit points (out of 100) on their course average to students who participated in the study. For example, if a student's course average was 90 (an A-) the student's final numerical grade was increased to 95 (an A).[8] Students who chose not to participate were allowed to do an extra credit project for the same five points.[9]

Fig. 1 depicts the flow of subjects in the experiment. Seven hundred and fifty-five students registered for the four sections of ECO 1001, of which 381 were in the Monday–Wednesday classes and 374 in the Tuesday–Thursday classes. Of the 755 registrants, 725 consented to be in the study, a 96% participation rate that represented 91% of all non-honors daytime students enrolled in ECO 1001. This participation rate is far greater than recent experimental studies of online learning. Thirty-two students either dropped the class before the midterm or did not take the midterm, and an additional 37 students took the midterm but afterwards either withdrew or did not take the final exam. The total post-randomization attrition rate was 9.5%.

We randomized students between formats within days (i.e. Monday–Wednesday or Tuesday–Thursday).[10] One section was taught in a large lecture hall that seats 274 students and the other section, taught at the same time, was in a classroom that held 114 students.[11] Each professor taught one compressed section and one traditional section, each in the same classroom. That is, Professor A taught a traditional section in the small classroom on Monday and Wednesday mornings and a compressed section in the same small room on Tuesday afternoons. Similarly, Professor B taught a traditional section in the large lecture hall on Tuesday and Thursday afternoons and the compressed section in the same room on Wednesday mornings. We can therefore control for the

[8] Approximately 132 students in each treatment arm were necessary for a minimum detectable effect size of 4.2 percentage points with 90% power. Lowering power to 80%, the required sample sizes fell to 98 students in each treatment arm. By offering an incentive to participate, we expected to recruit most of the 776 students that were likely to register for the four experimental sections, allowing more than sufficient remaining sample even if withdrawal had been substantially greater than the 10% observed in the study. These five extra credit points indeed proved crucial to recruitment. The IRB also allowed us to offer a raffle in which 40 students picked randomly from the participants would be given priority registration for their classes in the spring of 2014. Comments from students suggested that the number of priority registrations was too few to be a significant incentive, but that the five extra-credit points for one of the 8 classes that determines admission to the business school was highly valued.

[9] Of the 26 non-participants who finished the course (two others withdrew and two did not take the final), only 11 (42%) completed the extra credit project.

[10] Students that registered for a Monday–Wednesday section could not be randomized into Tuesday–Thursday sections because it would have potentially created conflicts with other classes for which they had registered.

[11] For administrative reasons, we were unable to secure two large lecture halls during the same class period for the experimental sections given existing accommodation of other large lecture classes.

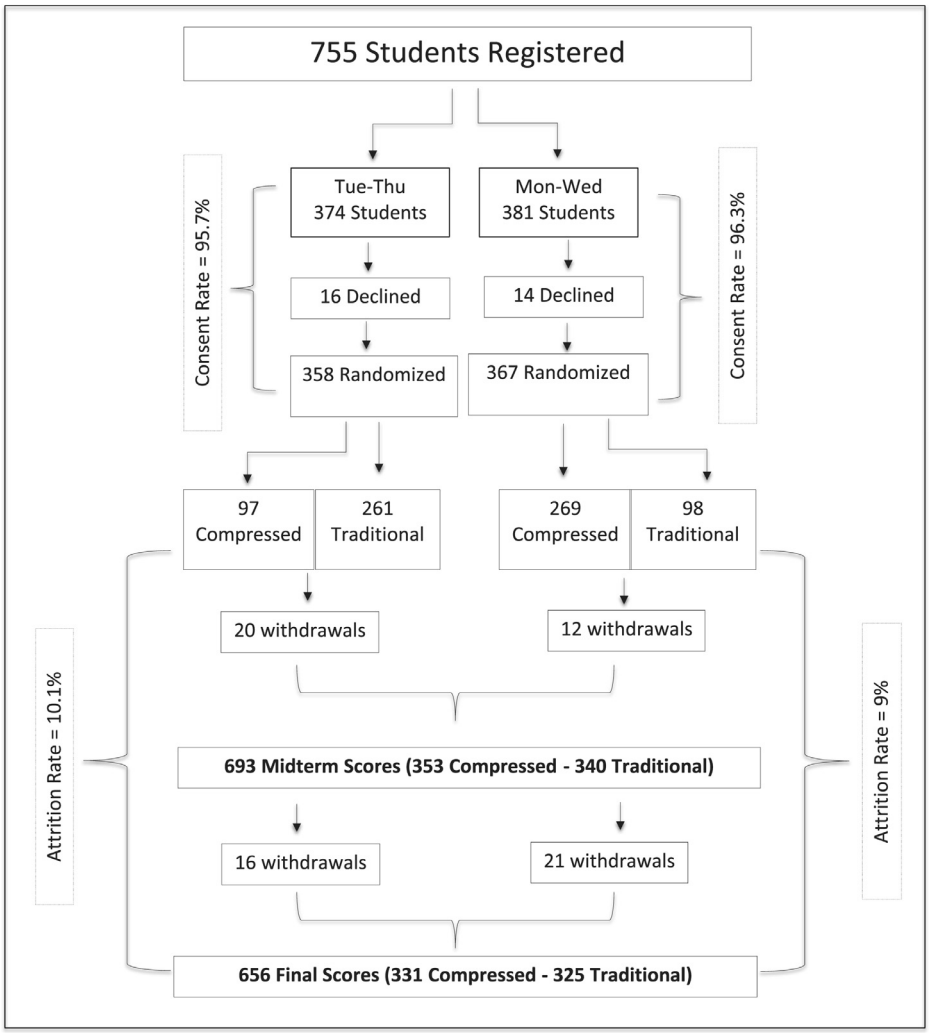

Fig. 1. Flowchart of student intake and random assignment.

professor and classroom size fixed effects, but we cannot separately identify them. Moreover, the administratively imposed restriction of having different class sizes introduces a potential source of treatment heterogeneity. "Within professor" comparisons contrast students from different randomized samples and "within day" comparisons contrast performance across classroom/professor. We present several sets of results: the pooled sample of all students with controls for day and classroom/professor, comparisons within classroom/professor, and third, comparisons within day but across classroom/professor and course format.

1.4. Outcomes

As with other experiments, our fundamental outcome measure is academic performance on exams and the final course grade. We administered both the midterm and final exams in class, and on both tests the same questions were used in all four sections. The midterm and final consisted of 30 and 40 multiple choice questions, respectively. The questions came both from a standardized test bank as well as being written by Professors A and B. A copy of each exam is included in Appendix A. We present results for the midterm and final separately as well as the total share of correctly answered questions on the combined midterm and final. We also present results with the overall course grade, in which the midterm and final exams counted for 35% and 45%, respectively. The remaining 20% of the course grade comprise online quizzes managed and graded by Aplia. The course grade also includes the penalty for missed classes described below, the five percentage-point bonus for participation, as well as curves for each exam.[12] In the results that we present below, we scale all test scores and the course grade so that they range from 0 to 100. We prefer the uncurved test scores as a measure of academic performance because, unlike the

[12] Each exam was curved so that the median curved exam score was 80%. As a result of this curve, 2 points (out of 30) were added to each midterm score and 6 points (out of 40) were added to each final exam score in the calculation of course grades.

course grade, it does not conflate non-cognitive (attendance) and cognitive (exams and online quizzes) outcomes.[13]

The primary purpose of the Aplia quizzes was to encourage students to keep up with the material and improve their preparation for the lecture. They were not supervised (i.e. we cannot determine whether students did their own work or worked with other students) and were intended as low-stakes assessments. Except for the week in which the midterm was given, students had a pre-lecture quiz that was graded on a pass-fail basis with only one attempt at the correct answer, and a post-lecture quiz that was graded on the percent correct.[14] Students were permitted three attempts at the correct answer on the post-lecture quiz and we used the average of all attempts. In calculating the contribution of the Aplia quizzes to the final grade calculation, each quiz was weighted proportionately to its total possible points (on average, the post-lecture quizzes were worth about three times as many points as the pre-lecture quizzes), and for each student we dropped the pre-lecture quiz and post-lecture quiz that most adversely affected his or her grade.

In addition to students' cognitive performance, we also examine whether the different formats elicited different amounts of non-cognitive effort. Within the first 15 min of each lecture students were required to swipe their student identification cards, giving us an accurate measure of attendance. Excluding the midterm and the first week of class, which did not count towards attendance requirements, students were allowed to miss 6 out of 25 lectures in the traditional format and three out of 12 in the compressed format without penalty, i.e. approximately 25 percent of the lectures. In the traditional classes, students lost one percentage point from their final grade for any late or missed classes beyond the six permitted absences, and in the compressed classes students lost two percentage points for any late or missed classes beyond the three permitted absences. The policy provided an incentive for students to swipe their ID cards, but it also created potentially meaningful variation in attendance within format.[15]

We also analyze withdrawal rates, counting as withdrawals students who enrolled in the class and consented to be in the study, but failed to finish.[16] Withdrawal rates are an important indicator of students' ability to manage a compressed format, but they also allow us to gauge the potential for attrition bias. Finally, we investigate online interaction by measuring how many course videos the students watched as well as the number of hours students spent logged in to Aplia.

2. Data

We combined several sources of data. All baseline characteristics were obtained from Baruch College's Office of Institutional Research and Program Assessment. These data included age, race/ethnicity, language spoken at home, major (if declared), grade point average (GPA), SAT scores, and cumulative credits. Some students have a GPA at Baruch, while transfer students have only GPA from their former college. Former transfer students have both GPAs. In the regression analysis that follows, we include both GPAs and indicator variables for missing one or both of those GPAs.[17] We also do not have SAT scores for all students because not all transfer students were required to submit their SAT scores to Baruch. We also administered two short surveys in the first and last week of classes, soliciting students' attitudes toward compressed courses and whether they held any employment during the semester.

3. Results

3.1. Summary statistics and balance

Table 1 contains baseline characteristics of students by format in the pooled sample. Characteristics of students at the start of the experiment are shown in the left panel and characteristics at the end of the semester are shown in the right panel. Overall there is strong balance, with no statistically significant differences between traditional and compressed formats on any of the individual baseline characteristics in the beginning sample and only one statistically significant difference (age) between the formats among students who took the final exam. For both samples we estimated a logit with an indicator for assignment into the compressed format as the dependent variable and the previously described student characteristics as the independent variables. The p-value for the overall χ^2 statistic from these regressions is 0.626 for the initial registrants and 0.157 in for the students who took the final exam. We also show the distribution of characteristics for the Monday–Wednesday and Tuesday–Thursday sections in Table A1 and these show similarly excellent balance.

Table 2 shows the baseline characteristics within professor/classroom. Estimating similar logit models as in Table A1 yields p-values that are larger than 0.05 for both the beginning and ending samples for Professor A and also in the beginning sample for Professor B, while in the ending sample for Professor B the differences are jointly significant at the 3.7% level. For both professors, we do find some

[13] Attendance is potentially endogenous and students could have worked with other students on their Aplia quizzes even for the questions that were algorithmically generated. Thus, the overall grade is a less-controlled measure of performance than the midterm and final exams.

[14] A student who answered at least 50% of the questions correctly on a pre-lecture quiz earned full points, while a student who answered less than 50% correctly received no adjustment. Thus 8 out of 15 correct was bumped up to 15/15, while 7 out of 15 was recorded as 7/15.

[15] Recitation sections, led by a graduate student, were held in conjunction with both large lectures. Each of the four recitations had a class size of almost 70 students. Attendance was voluntary, however, and extremely low. On average, students attended 1.1 recitations out of a possible 13 and the median and modal number of recitations attended was zero. There was no recitation available to students in the smaller classroom. Given the low participation rate, however, the presence of recitations should have little impact on the results.

[16] As noted, ECO 1001 is one of the eight classes that determine entrance to the Zicklin School of Business. Students can withdraw or not even show up for the final and accept a grade of F because they can retake the class and replace the F on their transcript. We treated official withdrawals and "no-shows" as the same. We also measured withdrawals between the midterm to the final.

[17] We have a GPA measure for about 78% of our sample. Baruch accepts many transfer students, particularly from other CUNY schools, and an additional 15% of the sample has information on their GPA at the school from where they transferred. About 20% of our sample has both a GPA measure from Baruch and from their previous institution.

Table 1
Baseline characteristics of participants at the beginning and end of the semester.

Covariate	Beginning sample				Ending sample			
	Traditional	Compressed	Compressed - traditional	*N*	Traditional	Compressed	Compressed - traditional	*N*
Prior academic performance								
Baruch GPA	3.00	3.01	0.01	568	3.01	3.06	0.05	518
Transfer GPA	3.31	3.26	−0.05	265	3.34	3.28	−0.06	230
SAT Verbal	541.56	533.31	−8.25	556	544.71	537.12	−7.60	511
SAT Math	601.90	596.17	−5.73	556	607.42	600.94	−6.48	511
Prior academic experience								
Cumulative Credits	45.93	44.98	−0.95	725	45.24	43.96	−1.28	656
Underclass	0.73	0.77	0.04	725	0.74	0.79	0.05	656
Attends Part Time	0.08	0.07	−0.00	725	0.08	0.07	−0.02	656
Demographic characteristics								
Age	21.22	20.93	−0.30	725	21.23	20.70	−0.53**	656
Female	0.45	0.48	0.02	725	0.44	0.46	0.02	656
Asian	0.44	0.43	−0.02	606	0.46	0.44	−0.03	546
Black, Hispanic, Other	0.31	0.28	−0.03	606	0.29	0.26	−0.03	546
Native English Speaker	0.54	0.53	−0.02	621	0.53	0.53	0.00	561
p-value, joint χ^2-test		0.626				0.157		

Note: Statistical significance tested using two-sample *t*-tests assuming unequal variances. Significance levels are indicated by * <0.10, ** <0.05, *** <0.01. The joint χ^2 tests are based on logit regressions of compressed on all variables shown in the table plus indicator variables for missing Baruch GPA, Transfer GPA, SAT scores, Race/Ethnicity, and Native English Speaker. Sample size for left panel is 725, sample size for right panel is 656.

differences in the proportion of Asian students, who were more likely to register for Tuesday–Thursday sessions than Monday–Wednesday sessions. There are also some statistically significant differences in prior academic experience for the students randomized into the sections taught by Professor A. Recall that we could not randomize within professor/classroom because we could not randomize across the Monday–Wednesday and Tuesday–Thursday schedules as this would have caused conflicts with students' other scheduled classes, and student preferences for taking classes on different days or at a different time of day may lead to some small differences between the compressed and traditional groups for each professor. Overall, however, the balance within professor is excellent, indicating that the randomization was successful.

3.2. Performance on tests and quizzes: pooled sample

We show differences across formats in student performance on the midterm, final, the combination of both, Aplia quizzes, and the final course grade in Table 3. As noted above, we scale all results to range from 0 to 100 to facilitate comparisons across the various performance metrics. For each outcome we show unadjusted (in odd-numbered columns) and adjusted (in even-numbered columns) mean percentage point differences. In all regressions in Table 3 we include an indicator for the Monday–Wednesday classes, although the coefficient on this variable is never statistically different from zero. Across all performance measures, we find that students in the compressed format did less well than students in the traditional format, and that these differences, except for Aplia scores, are statistically significantly different from zero. Adjusting for baseline covariates narrows the estimated mean differences between formats by a few tenths of one percentage point relative to the unadjusted differences. This similarity of the coefficient magnitudes in the unadjusted and adjusted specifications speaks to the balance in the pre-treatment covariates. The covariates also substantially increase the explanatory power of the model (increasing the R^2 from around 1% to 30% or more) and also yield moderate efficiency gains.

Students in the compressed format scored 3.2 percentage points less on the midterm or 0.21 standard deviations relative to those in the traditional format (columns 1 and 2). We also present separate results for the midterm for those students who completed the class in columns 3 and 4. The results are nearly identical to those in columns 1 and 2, suggesting that there is not selective attrition between the formats. This is confirmed in the results in Table 7 below, where we find no differences across format in the overall withdrawal rate or withdrawal after the midterm.

The differences between formats for the final exam, adjusted for covariates, are half as large in absolute value as those for the midterm and are statistically insignificant at conventional levels. [18] Results for the final suggest that students may have become more accustomed to the compressed format over the second half of the semester. The effect of format on the combined the scores for the midterm and final is 2.3 percentage points (p-value < 0.05) or 0.18 standard deviations. The lower bound of the 95 confidence interval of the estimated effect in column 8 is −4.2, or approximately one half of a letter grade. The results were nearly identical for the overall course grade (columns 11 and 12).[19]

In Fig. 2 we show kernel density estimates of all of outcomes in Table 3 for the compressed and traditional formats to examine visually differences in performance across

[18] The p-value is 0.14, with a 95% confidence interval of (−4.01, 0.52).

[19] Our results are similar to those of Figlio, Rush, and Yin (2013) who reported that students in the live lecture of introductory economics scored a statistically significant 2.5 percentage points higher on the average of three exams relative to those in the online section, adjusted for covariates. Our point estimates are also within the 95% confidence intervals of the difference between scores obtained by students in the hybrid and traditional statistics classes studied by Bowen et al. (2014).

Table 2
Baseline characteristics of participants at the beginning and end of the semester by professor/classroom size.

Covariate	Beginning sample				Ending sample			
	Traditional	Compressed	Compressed - traditional	*N*	Traditional	Compressed	Compressed - traditional	*N*
				Professor A/small classroom				
Prior academic performance								
Baruch GPA	3.06	2.89	−0.18	143	3.08	2.95	−0.13	131
Transfer GPA	3.37	3.32	−0.05	71	3.42	3.31	−0.11	65
SAT Verbal	543.52	520.67	−22.85	146	545.61	520.14	−25.47	138
SAT Math	609.01	594.53	−14.48	146	614.39	596.11	−18.28	138
Prior academic experience								
Cumulative Credits	48.80	42.58	−6.22*	195	47.94	41.87	−6.08*	181
Underclass	0.67	0.81	0.14**	195	0.70	0.82	0.12*	181
Part time	0.11	0.05	−0.06	195	0.12	0.04	−0.07*	181
Demographicw characteristics								
Age	21.26	20.89	−0.37	195	21.27	20.67	−0.60	181
Female	0.48	0.48	0.00	195	0.48	0.47	−0.01	181
Asian	0.36	0.58	0.22***	160	0.39	0.58	0.19**	148
Black, Hispanic, Other	0.30	0.18	−0.12*	160	0.26	0.17	−0.09	148
Native English Speaker	0.54	0.51	−0.03	171	0.52	0.50	−0.02	161
p-value, joint χ^2-test			0.126				0.221	
				Professor B/large classroom				
Prior academic performance								
Baruch GPA	2.98	3.05	0.08	425	2.98	3.10	0.11*	443
Transfer GPA	3.29	3.23	−0.06	194	3.31	3.26	−0.05	157
SAT Verbal	540.86	537.78	−3.08	410	544.38	543.38	−1.00	373
SAT Math	599.34	596.75	−2.60	410	604.83	602.72	−2.11	373
Prior academic experience								
Cumulative Credits	44.85	45.84	0.99	530	44.17	44.73	0.56	475
Underclass	0.75	0.75	0.01	530	0.75	0.78	0.03	475
Part time	0.07	0.08	0.02	530	0.07	0.07	0.01	475
Demographic characteristics								
Age	21.21	20.94	−0.27	530	21.22	20.71	−0.50*	475
Female	0.44	0.47	0.03	530	0.43	0.45	0.03	475
Asian	0.48	0.38	−0.10**	446	0.49	0.39	−0.10**	398
Black, Hispanic, Other	0.31	0.31	−0.00	446	0.29	0.30	0.01	398
Native English Speaker	0.54	0.53	−0.01	450	0.53	0.54	0.01	400
p-value, joint χ^2-test			0.167				0.038	

Note: Statistical significance means between traditional (lectures twice per week) and compressed (lectures once per week) tested using two-sample t-tests assuming unequal variances. Significance levels are indicated by * <0.10, ** <0.05, *** <0.01. The joint χ^2 tests are based on logit regressions of compressed on all variables shown in the table plus indicator variables for missing Baruch GPA, Transfer GPA, SAT scores, Race/Ethnicity, and Native English Speaker. Sample sizes are 195 (beginning) and 181 (ending) for the top panel and 530 (beginning) and 475 (ending) for the bottom panel.

Table 3
Student performance.

Covariate	Midterm, all		Midterm, finishers		Final		Midterm + final		Aplia quizzes		Course grade	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Compressed	−3.77***	−3.24***	−3.30**	−3.26***	−2.42*	−1.64	−2.80**	−2.33**	−0.99	−1.28	−2.86**	−2.59***
	(1.45)	(1.16)	(1.43)	(1.16)	(1.32)	(1.10)	(1.24)	(0.97)	(1.75)	(1.48)	(1.24)	(0.96)
Mon.-Wed.	<0.01	−1.09	0.23	−0.87	0.34	−1.02	0.29	−0.96	−1.50	−2.06	−0.03	−1.20
	(1.45)	(1.18)	(1.43)	(1.19)	(1.32)	(1.11)	(1.24)	(0.98)	(1.74)	(1.55)	(1.24)	(0.97)
Prof. A/Small Class		3.67***		2.78***		3.14***		2.98***		1.60		2.70***
		(1.14)		(1.14)		(1.10)		(0.95)		(1.56)		(0.96)
Other covariates		X		X		X		X		X		X
R^2	0.013	0.383	0.010	0.378	0.006	0.325	0.010	0.429	0.003	0.311	0.010	0.457
N	693		656		656		656		656		656	
Mean score, traditional	73.17		74.16		60.98		66.63		78.66		82.93	
Standard dev., traditional	15.54		14.93		14.85		13.08		19.20		13.11	

Note: All outcomes are based on a 100-point scale. Estimated with OLS. Heteroskedasticity-consistent standard errors in parentheses. Significance levels are indicated by * <0.10, ** <0.05, *** <0.01. Other covariates are Baruch GPA, Transfer, GPA, Verbal SAT, Math SAT, Cumulative Credits, Age, indicator variables for Part-Time Student, Underclassman, Female, Asian, Black/Hispanic/Other, and Native Speaker plus indicator variables for missing Baruch GPA, Transfer GPA, SAT scores, Race, and Native English Speaker. Mean scores are for students in the traditional format. Midterm, Final, and Midterm + final are raw (uncurved) scores. Aplia is average score on online quizzes. Course Grade includes curved midterm and final grades, penalties for missed classes, and the 5 percentage point participation bonus. For the complete regression results see Table A2 of the Appendix.

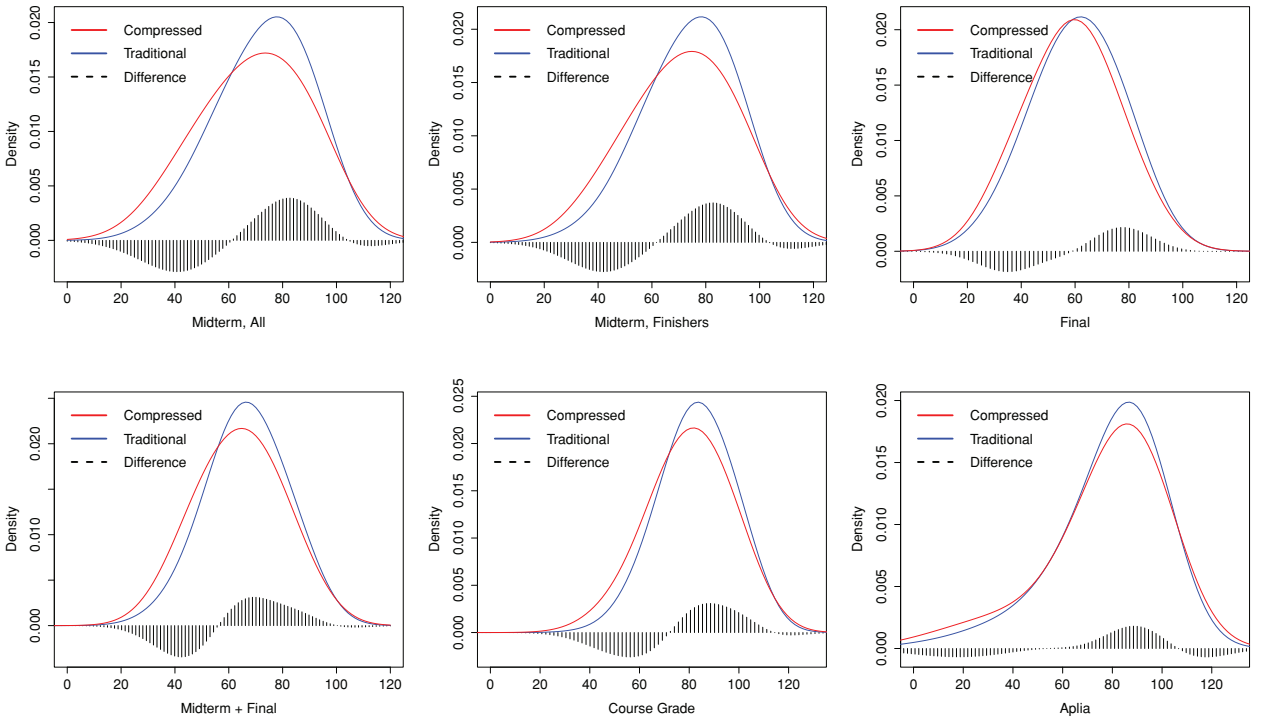

Fig. 2. Kernel density estimates of student performance. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

formats. The red lines indicate the densities for the distribution of outcomes in the compressed sections while the blue lines indicate the distribution of outcomes in the traditional sections. The shaded area below each plot shows the difference in densities between traditional and compressed formats. The plots reveal a roughly symmetrical distribution of exam scores and the final grade, with the distribution in the compressed format shifted slightly left to that of the traditional distribution. The exception is the distribution of scores on the Aplia quizzes, which are nearly identical across formats (though clearly skewed left, reflecting that students were allowed three attempts to answer post-lecture quizzes correctly and some students failed to submit several assignments). We performed two-sample Kolmogorov–Smirnoff tests on the difference between the densities in each of the panels of Fig. 2. We marginally rejected the null hypothesis of equal densities only for the midterm, where the test had a p-value of 0.078. We also performed two-sample Kolmogorov–Smirnoff tests on the difference in the raw (unsmoothed) distribution between the compressed and traditional sections for all of the outcomes in Table 3, and rejected the null hypothesis of equal distribution for the midterm (p-value of 0.04), midterm plus final (p-value of 0.02), and the course grade (p-value of 0.10) but not for the final exam (p-value of 0.23).

3.3. Performance on tests and quizzes within professor/classroom

As noted earlier, we were unable to procure classrooms of equal size for administrative reasons. Each professor taught only in either a small classroom with a capacity of 114 students (Professor A) or in a large classroom with a capacity of 274 students (Professor B). Although we include professor/classroom fixed effects in Table 3, we cannot separately control for or distinguish between the effects of heterogeneous professors and classroom sizes. To examine whether this is an issue, we present estimates of the treatment effects separately for each professor/classroom in Table 4. The top panel shows the results for Professor A (in the smaller lecture hall) while the bottom panel shows the results for Professor B (in the larger lecture hall). The outcomes are the same as in Table 3 and columns present unadjusted and adjusted treatment effects as in Table 3.

Overall, the point estimates are qualitatively consistent with those from the pooled sample shown in Table 3. Students in the compressed section taught in the large lecture hall (Professor B) scored approximately 4.5 percentage points lower on the midterm but 2.6 percentage points lower on the final. There are generally less differences by format in the smaller classroom with Professor A, but the standard errors are larger resulting in statistically insignificant differences for the combined midterm and final exams, the final exam, and the overall grade. The estimated differences are also more sensitive to the inclusion of covariates than those in the pooled sample in Table 3. Recalling that we only randomized within days, the within professor/classroom estimates therefore compare students from two different randomized samples. Although the balance of baseline characteristics by format appears reasonable, there are greater differences in some characteristics by format, as shown in Table 2. We view these results as comparable to those from the pooled sample, however, while eliminating an important source of heterogeneity.

Table 4
Student performance within professor/classroom.

Covariate	Midterm, all		Final		Midterm + final		Aplia quizzes		Course grade	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Professor A/small classroom										
Compressed	−3.71	−2.82	−2.69	−0.28	−3.01	−1.33	0.57	0.24	−2.76	−1.57
	(2.46)	(2.19)	(2.23)	(2.05)	(2.13)	(1.85)	(2.92)	(2.47)	(2.12)	(1.77)
Other covariates		X		X		X		X		X
R^2	0.012	0.460	0.008	0.415	0.011	0.490	<0.001	0.338	0.009	0.503
N	184		181		181		181		181	
Mean score, traditional.	76.16		63.61		69.13		79.34		85.11	
Std. dev., traditional	16.01		14.51		13.91		21.21		13.96	
Professor B/large classroom										
Compressed	−3.70**	−4.50***	−2.04	−2.62**	−2.47**	−3.39***	−2.47	−3.38**	−2.86**	−3.87***
	(1.48)	(1.22)	(1.37)	(1.22)	(1.23)	(1.00)	(1.90)	(1.64)	(1.27)	(1.00)
Other covariates		X		X		X		X		X
R^2	0.012	0.365	0.005	0.302	0.008	0.417	0.004	0.338	0.011	0.459
N	509		475		475		475		475	
Mean score, traditional	72.04		59.95		65.65		78.39		82.07	
Std. dev., traditional	14.51		14.89		12.64		18.39		12.69	

Note: All outcomes are based on a 100-point scale. Estimated with OLS. Heteroskedasticity-consistent standard errors in parentheses. Significance levels are indicated by * <0.10, ** <0.05, *** <0.01. Other covariates are Baruch GPA, Transfer, GPA, Verbal SAT, Math SAT, Cumulative Credits, Age, indicator variables for Part-Time Student, Underclassman, Female, Asian, Black/Hispanic/Other, and Native Speaker plus indicator variables for missing Baruch GPA, Transfer GPA, SAT scores, race/ethnicity, and Native English Speaker. Mean scores are for students in the traditional format. Midterm, Final, and Midterm + final are raw (uncurved) scores. Aplia is average score on online quizzes. Course Grade includes curved midterm and final grades, penalties for missed classes, and the 5 percentage point participation bonus. Capacity of the small classroom is 114 students while the large classroom is 274 students.

Table 5
Student performance within class day.

Covariate	Midterm, all		Final		Midterm + final		Aplia quizzes		Course grade	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Monday–Wednesday classes										
Compressed	−7.83***	−6.83***	−5.71***	−4.46***	−5.95***	−5.05***	−3.42	−2.51	−5.90***	−4.96***
	(2.00)	(1.57)	(1.79)	(1.40)	(1.71)	(1.25)	(2.65)	(2.24)	(1.74)	(1.25)
Other covariates		X		X		X		X		X
R^2	0.037	0.434	0.029	0.409	0.035	0.495	0.005	0.365	0.032	0.524
N	355		334		334		334		334	
Mean score, traditional	76.16		63.61		69.13		79.34		85.11	
Std. dev., traditional	16.01		14.51		13.91		21.21		13.96	
Tuesday–Thursday classes										
Compressed	0.42	1.07	0.98	1.62	0.47	0.85	1.53	1.22	0.28	0.49
	(2.05)	(1.78)	(1.91)	(1.78)	(1.76)	(1.55)	(2.25)	(2.05)	(1.75)	(1.53)
Other covariates		X		X		X		X		X
R^2	<0.001	0.340	0.001	0.298	<0.001	0.386	0.001	0.288	<0.001	0.404
N	338		322		322		322		322	
Mean score, traditional	72.04		59.95		65.65		78.39		82.07	
Std. dev., traditional	16.01		14.89		12.64		18.39		12.69	

Note: All outcomes are based on a 100-point scale. Estimated with OLS. Heteroskedasticity-consistent standard errors in parentheses. Significance levels are indicated by * <0.10, ** <0.05, *** <0.01. Other covariates are Baruch GPA, Transfer, GPA, Verbal SAT, Math SAT, Cumulative Credits, Age, indicator variables for Part-Time Student, Underclassman, Female, Asian, Black/Hispanic/Other, and Native Speaker plus indicator variables for missing Baruch GPA, Transfer GPA, SAT scores, Race, and Native English Speaker. Mean scores are for students in the traditional format. Midterm, Final, and Midterm+final are raw (uncurved) scores. Aplia is average score on online quizzes. Course Grade includes curved midterm and final grades, penalties for missed classes, and the 5 percentage point participation bonus.

3.4. Performance on tests and quizzes within day

To illustrate the importance of professor/classroom heterogeneity, we show estimates of the compressed class time effect comparing formats within day in Table 5. The top panel (Monday–Wednesday) compares outcomes of students in which the compressed format was delivered in the large lecture hall and the traditional format in the smaller room. In the bottom panel (Tuesday–Thursday) the opposite occurred: students in the compressed format were in the smaller classroom and those in traditional format had class in the large lecture hall. The differences are striking. Students in the compressed format scored over 5 percentage points less or 0.36 standard deviations on the combined midterm and final (top panel, column 7) than those in the traditional class when the compressed was delivered in the large lecture hall,

Table 6
Student performance stratified by tercile of predicted student performance.

Predicted performance	Midterm, all		Midterm, finishers		Final		Midterm + final	
	Coeff.	Mean	Coeff.	Mean	Coeff.	Mean	Coeff.	Mean
Low	−5.42**	63.5	−5.43***	65.3	0.10	51.5	−2.02	56.9
	(2.29)	(1.40)	(2.07)	(1.37)	(1.75)	(1.18)	(1.66)	(1.01)
Medium	−4.13**	74.5	−3.78**	76.0	−3.21*	62.4	−3.41**	68.4
	(1.80)	(1.17)	(1.75)	(1.12)	(1.70)	(1.25)	(1.54)	(0.94)
High	−0.68	83.1	−0.56	83.2	−2.30	70.8	−1.56	76.2
	(1.73)	(1.11)	(1.75)	(1.21)	(1.73)	(1.19)	(1.52)	(1.07)
N	693		656		656		656	

Note: All outcomes are based on a 100-point scale. Estimated with OLS using the repeated split sample (RSS) estimator in Abadie et al. (2014). Bootstrapped standard errors with 1000 replications in parentheses. Number of repeated split sample repetitions is 200. Significance levels are indicated by * <0.10, ** <0.05, *** <0.01. Covariates that are used to predict the student performance terciles are Baruch GPA, Transfer, GPA, Verbal SAT, Math SAT, Cumulative Credits, Age, indicator variables for Part-Time Student, Underclassman, Female, Asian, Black/Hispanic/Other, and Native Speaker plus indicator variables for missing Baruch GPA, Transfer GPA, SAT scores, Race, and Native English Speaker. All regressions control for professor/classroom fixed effects. Mean scores are for students in the traditional format. Midterm, Final, and Midterm + final are raw (uncurved) scores.

but there was no difference between formats when the compressed class was given in the smaller classroom (lower panel, column 7). That is, students that met once a week in the smaller classroom did as well as students that met twice a week in larger classroom; all of the estimated effects of being in the compressed format are positive (although statistically zero) in the Tuesday–Thursday sections. These differences are not likely due to imbalance between students in the two formats because randomization occurred within day (see Table A1) and are at least suggestive that smaller class sizes may play an important role in learning.[20]

3.5. Heterogeneity in performance on tests

If there is treatment heterogeneity, our previous results would mask differences in performance for students with different characteristics. One common approach to heterogeneity is to interact the treatment indicator with baseline characteristics such as race, sex, or, in our case, student GPA. However, such an approach can lead to specification searches. An alternative borrowed from the clinical literature is to use the controls and baseline characteristics to estimate a prediction equation for the outcome of interest. The estimated coefficients from this regression are used to predict the outcome for the full experimental sample. The predicted outcomes are stratified into quantiles and treatment effects are estimated within each quantile. Abadie, Chingos, and West (2014) show that this procedure leads to overfitting and serious bias in finite samples.[21] They propose, instead, that the predicted outcome for the controls is consistently estimated using a leave-one-out or a repeat split sampling procedure in the initial prediction regression. In Table 6 we apply their repeated split sampling algorithm to our sample by treating the students in the traditional format as controls and estimating treatment effects for the entire sample within terciles of predicted performance. The results in Table 6 indicate that students in the bottom and middle terciles of predicted performance scored 5.4 and 4.1 percentage points lower, respectively, in the compressed format on the midterm, but with no differences in the top tercile. On the final exam, however, there are no differences between the formats in the bottom and top terciles, while the differences in the middle tercile are only marginally significant. This pattern persists for the combined midterm and final. Unsurprisingly, high performing students did equally well in the two formats on all cognitive outcomes.[22]

3.6. Attendance, online usage, attrition, and other classes

In addition to test scores, we also examine the effect of being in the compressed format on a variety of non-cognitive outcomes related to effort. In Table 7 we present the impact of the compressed format on attendance, the number of videos watched, time spent online using Aplia, the probability of withdrawing from the class at any time, and the probability of withdrawing from the class after the midterm. Columns 1 and 2 show that there is no difference between the formats in the average proportion of classes attended. We do find, however, that students in the smaller classroom taught by Professor A were somewhat more likely to attend and that students that had the morning Monday–Wednesday lectures were 2.2 percentage points less likely to attend than those with in the late afternoon Tuesday–Thursday lectures. We should note, however, that 17 percent of students in the compressed class were penalized for excessive absences, relative to 9% in the traditional format.

In column 3 we show that students in the compressed format had 1.8 more video views than students in the traditional format relative to a mean of 8.5 views. It is noteworthy that when we add the professor/classroom fixed effect (and other covariates) in column 7, we find that students

[20] The term "smaller" is relative as class sizes of 100 might be viewed as large on many campuses. Nevertheless, there is more interaction between students and faculty in a classroom with 100 students compared to a lecture hall with 270 students.

[21] In previous high-profile social experiments, researchers have used baseline characteristics among the controls to predict and stratify treatment effects among all experimental participants. Abadie, Chingos, and West (2014) demonstrate that in finite samples, the procedure leads to an overestimation of treatment effects in the lower quantiles and an underestimation in the upper quantiles.

[22] We also interacted the hybrid indicator with individual covariates. No single covariate was statistically significant although the point estimates for students in the compressed format that worked 30 or more hours per week were negative and relatively large in absolute value (see Joyce et al., 2014).

Table 7
Attendance, attrition, and online usage.

Covariate	Percentage attended		Number of videos		Hours on Aplia		Withdrew any time		Withdrew after midterm	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Compressed	1.21	0.98	1.82	2.52*	0.45	0.29	0.007	0.005	0.016	0.009
	(1.11)	(1.11)	(1.76)	(1.50)	(2.48)	(2.19)	(0.023)	(0.023)	(0.015)	(0.016)
Monday–Wednesday	−1.91	−2.23**	−0.16	−0.65	2.91	3.34	−0.014	−0.015	0.004	0.012
	(1.11)	(1.12)	(1.75)	(1.56)	(2.47)	(2.32)	(0.023)	(0.023)	(0.015)	(0.014)
Professor A/small class		2.03*		8.58***		−2.07		−0.034		−0.053***
		(1.09)		(1.60)		(2.34)		(0.024)		(0.015)
Other covariates		X		X		X		X		X
R^2	0.004	0.104	0.003	0.177	0.003	0.182	<0.001	0.069	0.002	0.065
N	656		656		656		725		693	
Mean outcome, traditional	85.02		8.54		44.26		0.095		0.044	
Std. dev., traditional	12.46		12.46		27.64		0.293		0.206	

Note: Estimated with OLS. Heteroskedasticity-consistent standard errors in parentheses. Significance levels are indicated by * <0.10, ** <0.05, *** <0.01. Percentage Attended is on a 100-point scale. Other covariates are Baruch GPA, Transfer, GPA, Verbal SAT, Math SAT, Cumulative Credits, Age, indicator variables for Part-Time Student, Underclassman, Female, Asian, Black/Hispanic/Other, and Native Speaker plus indicator variables for missing Baruch GPA, Transfer GPA, SAT scores, Race, and Native English Speaker. Mean outcomes are for students in the traditional format. 69 students in total withdrew at any time during the course, and 27 withdraw after the midterm. Withdrawal after the midterm is conditional on having taken the midterm.

whose professor was in the videos watch 8.6 more videos than those whose professor was not in the videos. Students appear drawn to videos in which *their* professor appears. The finding argues for personalizing online material as much as possible. We find no differences in the number of hours students spent on Aplia (columns 5 and 6), although students spent a substantial amount of time on Aplia. The mean was 44 h or about 3.1 h per week over 14 weeks. Although the result is not statistically significant, there appears to be some evidence that students in Professor A's classes substituted time watching videos for time on Aplia. Overall, student effort as measured by attendance, videos, and online quizzes was largely the same by format. Importantly, students in the compressed format did not appear to substitute more use of the measurable online material for reduced time they spent in the classroom. While it is possible, of course, that students in the compressed format spent more time studying the textbook or with other online materials that we do not measure, we suspect that reducing time in class leads to a decrease in the total amount of time that students were engaged with ECO 1001 relative to those in the traditional format.

The lack of differences in attendance and intensity of online usage by format indicates that students in the compressed class had, on average, a minimum of 13.8 more hours during the semester to apply to other material related ECO 1001 or to their other courses.[23] As a check we tested for variation in student grades in the other classes taken in the same semester with ECO 1001 by format but found no differences across students in different formats of ECO 1001.[24] In columns 7 and 8 we present results indicating that students did not withdraw more at any time from the class in the compressed sections and in columns 9 and 10 we find the same result for withdrawal after the midterm. These findings indicate that attrition bias is unlikely to affect our results.

3.7. Student surveys: preference for compressed or traditional

We surveyed students in the first and last week of classes about their preferences regarding class formats. In the first week of class, we asked students to rate the statement, "I would have chosen the hybrid over the traditional format if I had had the choice," on a four-point Likert scale ranging from "Strongly Agree" to "Strongly Disagree".[25] The students exhibited a strong *a priori* preference for the compressed format, but it varied by their random (but known) assignment: 78% of those randomly assigned to the compressed format agreed or strongly agreed, but only 55% in the traditional format did. When we asked students at the end of the semester (but prior to the final exam or knowing their final grade) if they would choose the same format for their next economics class, the results shifted in favor of the traditional format: 65% of the students in the traditional class but only 54% in the compressed format agreed or strongly agreed. Thus, the preference for the traditional format increased by 20 percentage points from the beginning of the semester amongst those in the traditional format, while the preference for the compressed format decreased by 24 percentage points among those in the compressed format. Despite this change in preferences, 67% of students in the traditional format and 62% in the compressed format agreed or strongly agreed that having class twice per week helped their grade, but 62% students in the compressed format disagreed with the statement that the compressed format hurt their grade. Somewhat surprisingly, we found no differences in responses when we stratified the data by the students' baseline GPA. We interpret the survey results to mean that students found the compressed format appealing before having experienced it, but found it challenging during the semester. We thus expect that a substantial proportion of students would not opt for the compressed format for their next economics class.

[23] The difference in the median number of classes attended between the two formats was 11, and each class period was 75 min. This figure does not include time getting to and from class.

[24] These results are available from the authors by request.

[25] We referred to the compressed format as a "hybrid" format in the student surveys.

4. Conclusion

We found that students in a traditional lecture format of introductory microeconomics, with twice as much face-to-face instruction, performed better on the midterm than students in a compressed version of the same class. Although this difference in performance was reduced by half for the final (suggesting that students in the compressed format adjusted to the reduced class time of the compressed format), students in the traditional format performed better on the combined midterm and final as well as in the final course grade. Students whose baseline predicted performance was in the upper tercile experienced no difference between formats on either the midterm or the final, suggesting that high performing students can succeed with less class time. We also found no difference by format when students in the compressed class in a classroom of 114 seats were compared to students in the traditional format in a lecture hall that has 272 seats. Although we cannot separately identify the effect of instructor from class size, the similarity in the quality of the instructors suggests class size matters.

We have improved on the existing literature in several important dimensions. First, we had a 96% participation rate and an attrition rate of 10% that did not vary across experimental treatments, strongly supporting a claim to internal validity. Second, each of the two participating faculty taught one of each format, which allowed us to control for a potential source of heterogeneity. Third, all students had access to the same lecture notes and online materials, eliminating an artificial and arguably unenforceable restriction of access to online materials for students in the traditional class. Lastly, our large sample, with 725 students at the beginning of the experiment and 656 at completion, increased the precision of our estimates.

Fifty percent of participants in the study were transfer students to Baruch, 21% from community colleges within the City University of New York system—a population similar to that in Bowen et al. (2014).[26] Our results are also

[26] Indeed Baruch College was one of the six sites in the study by Bowen et al. (2014).

Table A1
Baseline characteristics of participants at the beginning and end of the semester by instruction day.

Covariate	Beginning sample				Ending sample			
	Traditional	Compressed	Compressed - traditional	*N*	Traditional	Compressed	Compressed - traditional	*N*
Monday–Wednesday								
Prior academic performance								
Baruch GPA	3.06	3.05	−0.01	286	3.08	3.10	0.02	260
Transfer GPA	3.37	3.23	−0.14	127	3.42	3.26	−0.16*	109
SAT Verbal	543.52	537.78	−5.74	283	545.61	543.38	−2.22	138
SAT Math	609.01	596.75	−12.27	283	614.39	602.72	−11.68	138
Prior academic experience								
Cumulative Credits	48.80	45.84	−2.96	367	47.94	44.73	−3.22	334
Underclass	0.67	0.75	0.08	367	0.70	0.78	0.08	334
Part time	0.11	0.08	−0.03	367	0.12	0.07	−0.05	334
Demographic characteristics								
Age	21.26	20.94	−0.33	367	21.27	20.71	−0.56	334
Female	0.48	0.47	−0.01	367	0.48	0.45	−0.02	334
Asian	0.36	0.38	0.01	315	0.39	0.39	−0.00	286
Black, Hispanic, Other	0.30	0.31	0.00	315	0.26	0.30	0.04	286
Native English Speaker	0.54	0.53	−0.01	317	0.52	0.54	0.02	286
p-value, joint χ^2-test			0.551				0.450	
Tuesday–Thursday								
Prior academic performance								
Baruch GPA	2.98	2.89	−0.09	282	2.98	2.95	−0.03	258
Transfer GPA	3.29	3.32	0.03	138	3.31	3.31	0.01	121
SAT Verbal	540.86	520.67	−20.19	273	544.38	520.14	−24.24*	250
SAT Math	599.34	594.53	−4.81	273	604.83	596.11	−8.72	250
Prior academic experience								
Cumulative Credits	44.85	42.58	−2.28	358	44.17	41.87	−2.30	322
Underclass	0.75	0.81	0.07	358	0.75	0.82	0.07	322
Part time	0.07	0.05	−0.01	358	0.07	0.04	−0.02	322
Demographic characteristics								
Age	21.21	20.89	−0.31	358	21.22	20.67	−0.54	475
Female	0.44	0.48	0.04	358	0.43	0.47	0.04	475
Asian	0.48	0.58	0.11	291	0.49	0.58	0.09	398
Black, Hispanic, Other	0.31	0.18	−0.13**	291	0.29	0.17	−0.12**	398
Native English Speaker	0.54	0.51	−0.03	304	0.53	0.50	−0.03	400
p-value, joint χ^2-test			0.366				0.080	

Note: Statistical significance means between traditional (lectures twice per week) and hybrid (lectures once per week) tested using two-sample *t*-tests assuming unequal variances. Significance levels are indicated by * <0.10, ** <0.05, *** <0.01. The joint χ^2 tests are based on logit regressions of compressed on all variables shown in the table plus indicator variables for missing Baruch GPA, Transfer GPA, SAT scores, race/ethnicity, and Native English Speaker. Sample sizes are 367 (beginning) and 334 (ending) for the top panel and 358 (beginning) and 322 (ending) for the bottom panel.

Table A2
Regression coefficients for student outcomes, Table 3.

Outcome: Table 3 Column:	Midterm, all (2)		Midterm, finishers (4)		Final (6)		Midterm + final (8)		Aplia quizzes (10)		Course grade (12)	
Covariate	Coeff	Std. Err.	Coeff	Std. Err.	Coeff	Std. Err.	Coeff	Std. Err.	Coeff	Std. Err.	Coeff	Std. Err.
Compressed	−3.238	1.159	−3.263	1.155	−1.636	1.101	−2.334	0.968	−1.276	1.476	−2.587	0.957
Monday–Wednesday	−1.088	1.181	−0.871	1.186	−1.019	1.114	−0.955	0.978	−2.056	1.546	−1.201	0.973
Professor A/small classroom	3.670	1.142	2.776	1.140	3.138	1.100	2.983	0.954	1.600	1.562	2.696	0.959
Verbal SAT/100	0.354	0.713	0.173	0.696	1.102	0.804	0.704	0.644	−1.830	1.011	0.169	0.632
Math SAT/100	6.494	0.803	6.148	0.824	3.896	0.859	4.861	0.721	2.949	1.162	4.442	0.722
Missing SAT scores	−4.035	1.861	−3.607	1.841	0.862	1.860	−1.053	1.596	0.796	2.261	−0.771	1.563
Baruch GPA	11.187	1.099	11.456	1.123	10.323	1.001	10.809	0.909	17.476	1.669	12.816	0.981
Missing Baruch GPA	2.188	1.551	2.996	1.568	2.238	1.603	2.563	1.357	4.885	1.869	3.275	1.315
Transfer GPA	9.542	2.441	7.895	2.388	7.257	2.613	7.530	2.126	8.665	2.713	7.662	1.910
Missing Transfer GPA	−0.517	1.728	−0.530	1.702	2.419	1.816	1.155	1.476	−2.086	2.467	0.364	1.482
Cumulative Credits	−0.023	0.044	−0.017	0.045	−0.040	0.042	−0.030	0.036	−0.229	0.062	−0.068	0.036
Underclassman	−1.796	2.116	−1.398	2.157	−2.355	2.121	−1.945	1.769	−6.786	2.765	−2.838	1.701
Part time	−2.793	2.569	−2.622	2.582	1.186	2.179	−0.446	2.099	−5.128	3.537	−1.220	2.176
Age	−0.372	0.262	−0.377	0.282	0.062	0.220	−0.126	0.214	0.436	0.268	−0.006	0.192
Female	−2.625	1.071	−3.166	1.053	−4.130	1.022	−3.717	0.862	−1.129	1.390	−3.112	0.852
Asian	−1.309	1.417	−0.988	1.410	−0.320	1.341	−0.606	1.119	0.431	1.897	−0.520	1.103
Black/Hispanic/other	−1.926	1.596	−1.614	1.565	−1.241	1.462	−1.401	1.252	−3.820	2.159	−2.283	1.279
Missing race	2.658	1.773	2.552	1.808	1.594	1.811	2.005	1.505	2.087	2.433	1.674	1.479
Native English Speaker	−0.070	1.223	0.059	1.207	−2.192	1.174	−1.227	1.003	−0.169	1.706	−1.088	1.008
Missing language	0.885	1.576	−0.431	1.612	−2.297	1.539	−1.498	1.277	−2.118	1.979	−1.646	1.245
Constant	−20.679	10.714	−12.576	10.455	−20.953	11.740	−17.363	9.671	−2.184	13.771	−1.834	9.043
R^2	0.383		0.378		0.325		0.429		0.311		0.457	
N	693		656		656		656		656		656	

Note: All outcomes are based on a 100-point scale. Estimated with OLS. Heteroskedasticity-consistent standard errors. Course grade includes curved midterm and final grades, penalties for missed classes, and the 5 percentage point participation bonus.

relevant to recent studies of online instruction at community colleges, because the vast majority of students at Baruch also commute (Jaggars & Xu, 2011; Xu & Jaggars, 2013).[27] Overall, we are most comfortable suggesting that our findings are likely to pertain to large urban public universities in which a substantial proportion of students commute and/or work.

Classroom space and professors' time are costly resources for any university. Administrators may look to shift some learning activities to less-costly alternatives like online platforms. Our results show that less time in the classroom may lead to lower academic performance for those students who are not near the top of the skill distribution. Whether students will adapt to new learning modalities that require them to do more learning outside of the classroom, and whether professors can optimize their teaching methods to help students achieve the same (or better) performance as they did in traditional lecture formats are open questions. Much more research is needed about how different in-class and out-of-class activities interact in producing academic achievement in higher education.

[27] Xu and Jaggars (2013) reported that community college students scored a full grade lower in courses delivered completely online relative to their counterparts who took courses in a traditional face-to-face environment. Differences in performance by format were much smaller in our study, which is further evidence that purely online courses may be more challenging for students that commute and/or work.

Appendix A.

See Tables A1, A2.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.econedurev.2015.02.007.

References

Abadie, A., M.M. Chingos, M.R. West. 2014. *Endogenous stratification in randomized experiments*. Manuscript, John F. Kennedy School of Government, Harvard University (http://www.hks.harvard.edu/fs/aabadie/stratification.pdf (last accessed February 2, 2015).

Bowen, W. G. (2013). *Higher education in the digital age*. Princeton, NJ: Princeton University Press.

Bowen, W. G., Chingos, M. M., Lack, K. A., & Nygren, T. I. (2014). Interactive learning online at public universities: Evidence from a six-campus randomized trial. *Journal of Public Policy Analysis and Management, 33*(1), 94–111.

Eurich, A. C. (1958). A foundation looks at college teaching by television. *Journal of Educational Sociology, 31*(9), 329–336.

Figlio, D., 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 Labor Economics, 31*(4), 763–784.

Jaggars, S.S. and Di X. 2011, March. *Online and compressed course enrollment and performance in Washington state community and technical colleges* (Community College Research Center Working Paper Number 31). Teachers College, Columbia University.

Joyce, T., Crockett, S., Jaeger, D.A., Altindag, O. and S.D. O'Connell. 2014, March *Does classroom time matter? A randomized field experiment of hybrid and traditional lecture formats in economics* (National Bureau of Economic Research Working Paper 20006).

Macmitchell, L. T. (1955). Are institutions planning ahead? *The Journal of Higher Education, 26*(9), 463–469+504.

Mankiw, N. G. (2012). *Principles of microeconomics* (6th). Mason, Ohio: South-Western.

Pennebaker, James W., Goslings, Samuel D., & Ferrell, Jason D. (2013). Daily online testing in large classes: Boosting college performance while reducing achievement gaps. *PLoS ONE, 8*(11), e79774. doi:10.1371/journal.pone.0079774.

Rudolf, F (1956). *Mark Hopkins and the log: Williams College 1836–1872*. New Haven, CT: Yale University Press.

Xu, D., & Jaggars, S. S. (2013). The impact of online learning on students' course outcomes: Evidence from a large community and technical college system. *Economics of Education Review, 37*, 46–57.