Do Students Know Best?

Choice, Classroom Time, and Academic Performance

Ted Joyce*
Baruch College, CUNY and NBER

Sean Crockett
Baruch College, CUNY

David A. Jaeger
CUNY Graduate Center, University of Cologne and NBER

Onur Altindag
CUNY Graduate Center

Stephen D. O'Connell
CUNY Graduate Center

Dahlia Remler
Baruch College, CUNY and NBER

October 2015

Acknowledgements: This research was supported, in part, by a CUNY Collaborative Incentive Research Grant 20 to Ted Joyce and David Jaeger. We thanks John Choonoo and Paul Bachler of Baruch's Institutional Research and Program Assessment for help with data. We received helpful comments from seminar participants Baruch College, Universidad Carlos III, Universitat Pompeu Fabra, the University of Michigan, the University of Reading, the CUNY Higher Education Policy Series and from Eric Chiang, our discussant at the American Economic Association's Conference on Teaching and Research in Economic Education in Minneapolis, MN, May 2015.

***Corresponding Author:** Baruch College, Department of Economics and Finance, 55 Lexington Avenue, New York, NY 10010. Email: Theodore.Joyce@baruch.cuny.edu

Do Students Know Best?

Choice, Classroom Time, and Academic Performance

Abstract

We compare student academic performance in traditional twice-a-week and compressed once-a-week lecture formats in introductory microeconomics between one semester in which students were randomly assigned into the formats and another semester when students were allowed to choose their format. In each semester we offered the same course with the sections taught at the same times in the same classrooms by the same professors using the same book, software and lecture slides. Our study design is modeled after a doubly randomized preference trial (DRPT), which provides insights regarding external validity beyond what is possible from a single randomized study. Our goal is to assess whether having a choice modifies the treatment effect of format. Students in the compressed format of the randomized arm of the study scored -0.19 standard deviations less on the combined midterm and final (p<.01) and -0.14 standard deviation less in choice arm (p<.01). There was little evidence of selection bias in choice of format. Future analyses of online learning formats employing randomization should consider DRPT designs to enhance the generalizability of results.

Just before the Internet became an important part of college instruction, David Romer (1993) asked whether undergraduate students in a large introductory course of economics should attend class. His results and those of subsequent papers strongly suggested that attendance improved academic results.[1] More than 20 years later, following the introduction of rich digital learning environments, we now question whether face-to-face class time is even necessary. While computers and the Internet have delivered a cornucopia of technological and pedagogical innovations, the challenge of obtaining estimates of the causal effects of these innovations on learning remains formidable. A student's decision to take classes online or in the classroom may be correlated with her motivation, conscientiousness, ability, and other time constraints – factors that also affect academic performance. Estimating how academic performance is affected by a student's choice of the location of learning (e.g. online, in the classroom, or a mixture of the two) entails overcoming the same sources of bias facing past researchers in their pursuit of causal estimates of attendance.

Recent studies of the effect of class format on student performance have attempted to overcome potential selection biases by randomizing students into purely online, partially online (i.e. "hybrid") and traditional lecture classes. Although randomized experiments are of enormous value in establishing causal effects, they are not without some shortcomings. First, students may decide not to participate in a randomized experiment to avoid the risk of being assigned to a class format that they do not prefer, undermining the generalizability of the experiment to the population of interest. Second, the very act of being randomized to a class format rather than choosing it could affect a student's motivation and other psychological factors that also potentially affect outcomes (Marcus et al. 2012). Third, outside of an experimental context,

[1] See Durden and Ellis (1995), Devadoss and Foltz (1996), Marburger (2001), Cohn and Johnson (2003), Kirby and McElroy (2003), Marburger (2006), Stanca (2006), and Dobkin, Gil and Marion (2010).

students choose their class format and may do so to match their learning styles to the format that most improves, or at least does not harm, their academic performance. Finally, the time, expense and difficulty of randomized experiments limit their number.

These challenges have led to the development of the doubly randomized preference trial (DRPT), in which subjects are first randomized into an experimental setting and a choice setting (Janevic et al., 2003; Long, Little and Lin, 2008; Little, Long and Lin, 2008; Shadish, et al., 2008). In a DRPT, those in the experimental setting are then randomized again between treatments while those in the choice setting select between treatments. The DRPT has many advantages. Estimates from the choice and experimental arms, separately and together, all generalize to the same population. The DRPT is also a rigorous "within-study" design that enables estimation of the selection bias when subjects choose their treatment and potentially identifies which control variables can reduce the selection bias (LaLonde 1986; Shadish, et al. 2008). Finally, DRPTs have been used to estimate how choice of treatment alters treatment effects. This is an important advantage of a DRPT over a randomized controlled trial (RCT) when subjects are not blinded to which treatment they receive — a common feature of most social science experiments.

Despite the proliferation of randomized designs in the social sciences, few studies have used a DRPT. One reason may be the expense of recruiting more subjects in order to power both the random and preference arms of the trial. Another drawback, as in any randomized trial, is that subjects may not participate in the study because they still risk being assigned to a non-preferred treatment. Even with these limitations, a DRPT can enhance substantially the external validity of an intervention by estimating the effect of treatment choices on outcomes.

In this study we use a design similar to a DRPT to analyze whether students′ choice of class format alters the estimate of class format on academic performance. We implemented a randomized experiment and an observational choice study in successive academic years, where all features were identical other than that we randomized in the first year and allowed students to choose their format in the second. In the Fall of 2013, we randomized 725 students between a traditional undergraduate lecture class of introductory microeconomics that met twice a week (150 minutes) and a compressed format that met only once a week (75 minutes). Students in both formats had access to the same lecture slides, online software, and videos. Joyce et al. (2015) discusses the results of the randomized experiment in detail. In the fall of 2014, we offered the same course taught at the same times, in the same classrooms, by the same professors using the same book, software and lecture slides. We let the 769 students enrolled in the course choose between a traditional and compressed format, however, and had them complete a survey instrument to measure characteristics that could drive their choice of format, such as hours worked and learning style. Finally, we compared the student performance on tests by format and year. We contend that the design approximates a DRPT because the Principles of Microeconomics is required early in a student′s course of study, resulting in a distribution of students between years ″as if random″ (Shadish, Cook, and Campbell pp. 302-3; Dunning pp. 235-39).

Our study has advantages over prior DRPTs despite the lack of explicit randomization between the experimental and choice arms. In contrast to Shadish, Clark, and Steiner (2008), we examine a high stakes course in a real world setting over an entire semester. Unlike Janevic et al.′s (2003) medical DRPT in which 37% of eligible women agreed to participate, 96% of students enrolled in the sections in our study participated in the randomized experiment (Fall

2013) and all students in the choice arm (Fall 2014) were included in the analysis. We therefore have almost no issues of self-selection into the study and our results generalize to the population of interest.

In our previous randomized experiment (Joyce et al., 2015), class time improved performance, but less so in an Internet-rich environment than earlier studies of attendance suggested. In the choice arm of the study, we find that class time improved performance even less than in the randomized arm. Differences in test scores between the traditional and compressed formats when student chose their format are roughly 0.10 standard deviations less ($p<0.30$) in favor of the traditional class than when students were randomized. We can interpret the difference in treatment effects between the randomized and choice arms as the modifying effect of treatment choice under the assumption that our controls remove all omitted variables bias associated with choice of treatment.[2] Factors that predict choice of format, such as preferences for teacher and student interaction have little predictive power of student performance, and strong correlates of student performance, such as grade point average, do not predict students′ choice of format.

Two broad lessons emerge from our results. First, observational studies of the effect of class time format may be subject to less bias than previously thought in large required introductory classes, although this issue must be examined in a variety of settings. Second, our attempt to approximate a DRPT provides a model for evaluating different learning formats in higher education. DRPTs allow researchers to establish the internal validity of an intervention′s estimate as well as its generalizability to non-randomized settings. This can make subsequent

[2] This assumption is also known as “ignorability,” “no remaining confounding,” and “no common causes not controlled for. “

observational studies more credible while avoiding the time, difficulty and expense of additional randomized experiments.

I. Previous Experimental Estimates

The most effective means of eliminating selection bias is by randomizing students between formats, although executing a randomized design on campuses can be quite challenging (Bowen et al. 2014).[3] Figlio, Rush, and Yin (2013) compared students who took introductory economics online versus a traditional lecture format at a major research university and found no mean differences in test scores between formats. Their results were sensitive to the inclusion of covariates, however. Students in the "live" format scored 3 points higher (out of 100) on the final and 2.5 points higher on the average of all three exams than students restricted to the video-taped lectures. They also found that performance deficits in the online class were greater for Hispanic and students with below median GPAs. Alpert, Couch and Harmon (2015) also randomized students taking introductory microeconomics at a large state university. They contrasted student performance on a cumulative final exam across three formats: traditional face-to-face classes, blended or hybrid classes and classes delivered completely online. They found no difference between the traditional and the blended class but a roughly four percentage point deficit for students in the online course. A third study with a randomized design, Bowen et al. (2014), compared students in a hybrid versus traditional class of introductory statistics across six public

[3] Researchers who have studied the effect of online formats on college student performance have used statistical approaches to adjust for selection bias such as propensity score matching and control function methods (Coates, et al. 2004; Anstine and Skidmore 2005; Gratton-Lavoie and Stanley 2009; Olitsky and Cosgrove 2014), instrumental variables (Xu and Jaggers 2013; Bettinger et al. 2014), or fixed effects at the student, class or instructor level with large state-wide databases (Xu and Jaggers 2013, 2014; Hart, Friedman and Hill 2014; Johnson and Mejia 2014). Our focus here is those studies using randomized designs that are most comparable to ours.

universities. As in previous studies, they found no statistically significant differences between formats on pass rates, final exam scores, and a standardized test of statistical literacy.

Joyce et al. (2015) used a randomized design to test whether students in a compressed format of introductory economics performed as well as students in a traditional lecture format.[4] Students in the traditional class scored 3.3 points higher (out of 100 or 0.21 standard deviations) on the midterm ($p<0.01$) but only 1.6 points more or 0.11 standard deviations on the final ($p<0.14$) relative to students in the compressed format. There were no differences in attrition or online usage by format.

Although each of the four studies had strong claims to internal validity, whether they are generalizable to other populations is unknown. In both Figlio, et al. (2013) and Bowen et al. (2014), less than 25 percent of eligible students agreed to participate, while in Alpert, Couch and Harmon (2015) 46 percent of students in an online class who enrolled in the experiment did not take the final as compared to 30 and 36 percent in the face-to-face and blended formats, respectively.[5] It is unclear why students chose not to participate in the studies or to complete the class once enrolled. For example, did students′ preferences for a traditional format prevent risk-averse students from agreeing to randomization? Is motivation to learn affected by assignment to a non-preferred treatment arm? How might these factors affect estimates of student performance in online formats when students choose a preferred style of class?

[4] This study is the experimental arm of the present analysis.

[5] Alpert, Couch and Harmon (2015) had 323 students in the study collected over four semesters. But they do not report how many students take principles of microeconomics overall. Given the size of the university, the number must be over one thousand.

II. Empirical Framework

In a DRPT, eligible subjects who consent to the study are randomized between an experimental arm and a choice arm. In the experimental arm subjects are randomized to the treatments while in the choice arm subjects select their treatment. The DRPT has its origins in the clinical literature in response to concerns about the generalizability of results from randomized trials. One concern was the effect of treatment preferences on outcomes when subjects were not blinded to their assignment. Subjects who are disappointed with their assignment or those who are enthusiastic may behave in ways that alter the results.[6] A second concern was that in many real-world settings, subjects choose their treatment, which may also affect their response to treatment (Rucker 1989; Brewen and Bradley 1989; Wennberg et al. 1993; Awad et al. 2000). For example, in the Women Take Pride (WTP) study, Janevic et al. (2003) used a DRPT to assess effectiveness of programs to enhance women's management of heart disease. They found that adherence was greater in the choice arm, although it had no effect on health. Long, Little, and Lin (2008) re-analyzed the data from the WTP study and described the assumptions necessary to identify preference effects in a DRPT. They defined preference effects as the difference in treatment effects within preference strata (i.e., the causal effect of treatment among those who prefer treatment minus the causal effect of treatment among those who prefer the control, a double difference). They found that adherence was strongly affected by treatment preference but a sickness index only modestly so.

Shadish, Clark and Steiner (2008) used a DRPT as a rigorous form of a within-study test of selection bias as first presented by Lalonde (1986). In the canonical application of a within-

---

[6] Cook and Campbell (1979) refer to "resentful demoralization" among subjects assigned to a non-preferred treatment.

study design, a comparison group drawn from extant data representative of the population under consideration is matched to the control group from the experiment. Because neither group has been exposed to the intervention, differences in outcomes are attributed to selection bias. Shadish, Clark and Steiner (2008) argue, however, that this form of within-study design confounds selection bias with differences in control group populations and also potentially with differences in measurement of outcomes and control variables (see also Heckman, Ichimura, Smith and Todd 1998; Cook, Shadish and Wong 2008). A DRPT isolates the effect of selection bias by construction, as randomization insures that subjects in the choice arm are drawn from the same population as subjects in the randomized arm with the same measurement of all outcomes and control variables. Shadish, Clark and Steiner's (2008) DRPT examined the effect of math versus vocabulary training on test performance among undergraduate psychology students. They found that their extensive set of covariates eliminated treatment selection bias.

The assumptions underlying a DRPT are best illustrated within a potential outcomes framework (Long, Little and Lin 2008; Marcus et al. 2012). Consider a student drawn from an eligible population who has consented to participate in a study in which she will be randomized to one of two arms, an experimental arm and a choice arm. If she is assigned to the experimental arm she will be randomized again to one of two treatments. If she is randomized to the choice arm she will choose one of the two treatments. In our context, the two treatments are a compressed class format that meets once per week and a traditional class format that meets twice per week. The student therefore has four potential outcomes: let P_{fd} be the potential performance of the student in a course by format (f=1 is compressed, f=0 is traditional) and design (d=1 is randomized, d=0 is choice). Because of the double randomization in the experimental arm we can express the potential outcomes as

$$E(P_{11} \mid f=1, d=1) = E(P_{11}) \text{ and } E(P_{01} \mid f=0, d=1) = E(P_{01}) \tag{1}$$

The treatment effect of format in the randomized design is therefore

$$\delta_r = E(P_{11} - P_{01}) . \tag{2}$$

Equivalent expectations for students in the choice arm yields

$$E(P_{10} \mid f=1, d=0) = E(P_{10}) \text{ and } E(P_{00} \mid f=0, d=0) = E(P_{00}) \tag{3}$$

giving the treatment effect in the choice arm:

$$\delta_c = E(P_{10} - P_{00}) \tag{4}$$

The potential outcomes in the choice arm are unlikely to be independent of the choice of format, however. Let x be a vector of student characteristics prior to the course. If we assume that potential outcomes in the choice arm are independent of the design due to randomization and independent of format conditional on x, then we can write the following:

$$E_x[E(P_{10} \mid f=1, d=0, x)] = E_x[E(P_{10} \mid x)] \tag{5}$$

and

$$E_x[E(P_{00} \mid f=0, d=0, x)] = E_x[E(P_{00} \mid x)] \tag{6}$$

allowing us to rewrite the treatment effect in the choice arm as

$$\delta_c = E_x[E(P_{10} - P_{00}) \mid x] \tag{7}$$

If the variables in x capture all of the determinants of choice that are also related to outcomes, then any difference between the estimated δ_r and δ_c represents the pure the effect of getting one's choice of treatment on student performance. "Choice" embodies a number of concepts that cannot separately be identified. Students may choose a format because more class time helps their learning or because less class time frees them to study more effectively on their own. If students choose the format that they believe will improve their performance, then we would expect to see smaller effects in the choice arm rather in the randomized arm. Students may choose a format for other reasons, such as convenience, however, which could increase the treatment effect of format between the two designs.

Regression specification

We use linear regression to estimate whether the choice of format alters the treatment effect of format on student performance. We assume that students randomized in the fall of 2013 and students who choose their format in 2014 are balanced along all observed and unobserved characteristics, i.e. that they were as good as randomly assigned between years. We show in Table 1 below that our samples are remarkably balanced in the baseline characteristics of the students in both academic years and will describe the institutional features justifying our assumption that the unobservables are also likely balanced across years. We estimate the following model:

$$P_{idf} = \alpha_0 + \alpha_1 C_{if} + \alpha_2 D_{id} + \alpha_3 (C_{if} \times D_{id}) + \sum \beta_k X_{kidf} + \epsilon_{idf} \tag{8}$$

where P_{idf} is the academic performance of student *i*, in design *d*, and format *f*. Let C_{if} be one if the student is in the compressed format and 0 if in the traditional format. Let D_{id} be one if the student is in the randomized arm and zero if she is in the choice arm and let X_{kidf} be a vector of covariates. A number of estimates are relevant from equation (5). First, $\alpha_1 + \alpha_3$ is the difference in performance between the compressed and traditional formats in the randomized arm (2013), whereas α_1 is the same difference in the choice arm (2014). Thus, α_3 is the difference-in-differences or the treatment effect of being in the compressed format relative to the traditional format in the randomized relative to the choice arm. We know from Joyce et al. (2015) that students in the compressed format of the randomized design had lower test scores than students in the traditional format. If choosing one's format improves performance then α_3 should be negative.

Setting and course

As noted previously, we randomized 725 students in the fall of 2013 between a traditional lecture format of introductory microeconomics that met twice a week for 150 minutes and a hybrid or compressed format that met once a week for 75 minutes. In the fall of 2014 we offered the exact same course taught by the same professors at the same times and in the same classrooms. Instead of randomizing students between formats, however, 769 students enrolled in the format of their choice.[7] As in the randomized design, both professors were listed as the course instructors so that students' choices were based on format and schedule.

A detailed description of the course is presented in Joyce et al. (2015); here we briefly summarize key aspects. Principles of Microeconomics (ECO 1001) is a required course for all students applying to the business program at Baruch College's Zicklin School of Business.[8] It also fulfills a social science requirement for non-business majors. Nearly one thousand students take ECO 1001 each fall. The City University of New York (CUNY) registrar listed the classes as traditional and hybrid. The CUNY registrar defines a course as a hybrid if between 30 to 79 percent of course material that would have been presented face-to-face is instead presented online.

[7] Eighty-nine students in the choice arm of the study were not enrolled in their preferred format base on the pre-course survey because of scheduling or lack of availability.. We report results dropping those students below.

[8] Baruch College, part of the City University of New York (CUNY) 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. 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.

In both 2013 and 2014 all sections used N. Gregory Mankiw's *Principles of Microeconomics* (6th Edition) as the textbook, along with Cengage Learning's Aplia web application to administer and grade online quizzes. Each week students in both the traditional and compressed format 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 which were graded as a percentage of 100. The midterm and final accounted for 35 and 45 percent of their course grade, respectively. Grades on the Aplia quizzes accounted for the 20 percent.

Our analysis hinges on the year in which students take the course as being as good as randomly assigned. Table 1 presents the baseline characteristics of students in both years of the study – they are comparable in every way. Our identification strategy also depends on the unobservable characteristics being comparable across years and there are several institutional reasons we expect that this is so. Economics 1001 is required for applying to the business school (and some other majors) at Baruch. Students have no non-honors, daytime alternatives to the sections used in the study during both years. Moreover, students cannot postpone taking the course without educational and possibly even financial consequences.

Data

Our primary outcome measures are 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 within each semester. The midterm and final consisted of 30 and 40 multiple choice questions, respectively. We attempted to keep the content of exam questions and their difficulty as similar as possible between the two years but to control for any (small) differences, we standardize scores on all tests to have a mean of zero and standard deviation of one. We also analyzed withdrawal rates, counting as withdrawals students who enrolled in the class but failed to finish.

We obtained student characteristics prior to enrollment in the course 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 GPAs 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.[9] We also do not have SAT scores for all students because not all transfer students were required to submit their SAT scores to Baruch during the admissions process.

Rubin (2007, 2008) and Cook, Shadish and Wong (2008) advocate estimating causal effects in observational studies by prospectively investigating and measuring all possible factors driving treatment selection so that they can be used as control variables (through propensity scores or other means). We used evidence in other settings (Shadish, Clark, and Steiner, 2008), instructor familiarity with students and the course, and informal interviews with past ECO 1001 students to hypothesize that student choice of compressed or traditional format might be affected by a variety of time constraint, preference, and learning style factors: commuting time, hours of work; general in-person vs. online/electronic orientation; general in-person vs. online/electronic learning style; risk aversion; experience with online format; and preference for quantitative

[9] 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.

courses. We developed a detailed questionnaire to measure these characteristics and administered it within the first week of class in 2014. A copy of the survey instrument is presented in the Appendix. Students received 3-extra-credit points added to the final course average if they completed both the pre-course and post-course surveys on time and 1.5 credits if they completed just one. CUNY's Institutional Review Board also required that students who did not want to participate in the survey be provided an alternative assignment for the same credit. Of 676 students who completed the class 648 (95.9%) competed the pre-course survey.

During the experimental year, a more limited pre-survey was administered, which included working hours. Crucially, students were also asked which format they preferred but only during the first week of classes and after they had been assigned to a format. Since students were randomized, their preferred format should have been balanced between those in the compressed and traditional formats; it was not, however, suggesting that students' reported preferences were altered by their assignment. Such effects are consistent with endowment effects in the psychology literature (e.g., Kahneman, Knetch and Thaler 1991). Nonetheless, we still have a measure of format preference in the experimental year and have the same measure in the choice year. While most students obtained their preferred format in the choice year, some students were unable to obtain their preferred format, or their preferred format at their preferred time slot.

III. Results

Analyzing balance by year and format

Characteristics of students who finished the course in the randomized arm in 2013 and the choice arm in 2014 are displayed in Table 1. We show the absolute differences, normalized differences and the log of the ratio of standard deviations between years. Based on a t-test, only the only statistically significant difference is the verbal SAT scores. Neither the normalized differences or the ratio of the standard deviations, however, suggest the difference in verbal SAT scores by year will cause important imbalance in the regressions. For the rest of the variables, the assessment of balance reveals no substantive concerns. Also of note is the lack of a difference in withdrawal rates. In each year between 10 to 12 percent of students withdraw, which is similar to the withdrawal rate in 2013 (Joyce et al. 2015). Moreover, the characteristics of students by format remain balanced if we include those who withdrew.[10] Thus, we have support, based on observables, that students in our study appear as if they were randomly distributed between the two years (study arms), a key assumption of a DRPT.

We also compare differences by format within each year in Table 2. As expected, characteristics are balanced by format in 2013 when students were randomly assigned. There is also no evidence of any meaningful imbalance by format along those observable characteristics in 2014 when students chose their format. Clearly, from this evidence, we do not know if there is imbalance in *unobservables* in the choice arm of 2014. We do know, however, that a) both observed and unobserved characteristics are balanced in 2013; b) there is balance in observables between 2013 and 2014; and c) there is balance by format in the same variables within the choice arm in 2014. We therefore have plausible support for a key assumption necessary to identify choice effects: that conditional on covariates, treatment effects in the randomized and choice arms are the same.

---

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

Regression estimates

Table 3 shows the regression results for the midterm, final, the combined midterm and final, the course grade and the probability of withdrawal. For each we show the effect of the compressed class relative to the traditional class in 2013 and 2014. The odd columns include a limited set of covariates and the even columns include a larger set. Coefficients show differences in standard deviations of standardized test scores and grades and differences in the probability of withdrawal. Students in the compressed format in 2013 scored between -0.14 and -0.22 standard deviations less in tests and their final grade. Format had no effect on differential withdrawal rates. In 2014, when students chose their format, those in the compressed format also scored lower, between -0.13 and -0.16 standard deviations, than in students in the traditional format.

Our test of whether student choice of format in 2014 modified the treatment effect was the difference in estimates of the compressed format in 2013 versus 2014. All differences were negative indicating that a positive effect of class time on academic performance decreases when students choose a format. However, differences were small. They ranged from -0.01 to -0.08 standard deviations across all measures of academic performance and none was statistically significant. Put differently, the option to choose a class format did not substantively mitigate the loss of class time on performance.

Estimates within professor/classroom

In Table 4 we show separate estimates within professor/classroom. The essential inferences form Table 3 persist but some differences are notable. First, there is no evidence of selection bias within each professor/classroom. The performance of students who chose the traditional format was no different from those who were randomized into it. This is evidence that a key assumption of the DRPT holds and that comparison of treatment effects between the randomized and choice arms of the study reflect the effect of choice of format on performance. In the larger classroom taught by professor B we find that the effect of the compressed format when students choose is approximately half as large as when students are randomized. Differences range from -0.06 to -0.13 standard deviations. These are not statistically significant from the estimates in the randomized design, but they are consistently smaller in magnitude suggesting that choice of format improved performance. The same pattern exists in the smaller classroom with professor A, although standard errors are substantially larger and estimates are more sensitive to adjustment.

Robustness checks

Eighty-nine students in the choice arm of the study were not enrolled in their preferred format base on the pre-course survey. Enrollment in the compressed class was full whereas enrollment in the traditional class was not, which suggests that the 89 students were constrained from taking the compressed format. We dropped these 89 students and re-estimated the regressions in Table 3. The results were essentially the same. For example, the difference in student performance on the combined midterm and final between the compressed and traditional formats in the randomized versus the choice arm is -0.06 (-0.05 in the full sample) with a standard error of 0.09. None of the other results differed meaningfully when the 89 students were dropped. Another issue is that the professors may have become more effective in delivering the compressed class in the second year relative to the first and it is possible that greater instructor experience and not choice of format may account for the relative improvement in test scores, albeit small, in the compressed format of the choice arm. We find no evidence to support this, however. In results available from the authors by request,, the effect of being in the compressed format did not differ between years.

Determinants of format

Results of the pre-class survey administered to students in the choice arm are shown in Table 5. As indicated in column (3), most questions used a 5-point Likert scale ranging from 1 (strongly agree) to 5 (strongly disagree) with 3 representing neutrality (neither agree or disagree). Treating the Likert responses as quantitative variables, columns (1) and (2) give the average score for each question for students in the traditional and compressed sections, respectively. Column (4) shows the chi-square test of independence among response categories for the full 2 (format) by 5 (scale) table for each question. Relatively few questions are associated with choice of format. Those who agreed that traditional lectures worked well for them were more likely to choose the traditional format, while those who agreed that their learning style was well-suited to a hybrid format were more likely to choose the compressed class. Those who preferred quantitative to writing-focused courses were more likely to choose the compressed format. Finally, those who preferred interaction with the professor and other students were more likely to choose the traditional format. Other factors such as commuting time or paid employment were not associated with choice of format.

In the first two columns of Table 6 we present results of a regression of student choice of format on student characteristics and the five survey questions that showed differences between formats from Table 5 ("learning style is well-suited to a hybrid format," "traditional lectures work for me," "I need structure to get my class work done," "Prefer quantitative courses to writing-focused courses," and "Interaction with professors and other students helps.") For the survey questions, we grouped the first two categories (strongly agree and agree) as well as the last two (disagree and strongly disagree); the latter is the reference category for each of the survey questions. None of the student characteristics, including those that are strong predictors of student performance like GPA predict format choice and of the survey questions being "neutral" about interaction with professors predicts format choice.

In columns (3) and (4) we show results for the effect being in the compressed format in 2014 along with the student characteristics on student performance on the combined midterm and final, essentially replicating the results from columns (5) and (6) of Table 3.[11] GPA, math ability as measured by the SAT, and the female dummy are all strong predictors of performance. In columns (5) and (6) of Table 6 we add the four survey questions to examine whether adding these indicators affect the estimated effect of being in the compressed sections. When we add these variables, the estimated treatment effect decreases by .033, or roughly 20 percent of the estimated effect without these variables, and is closer to the experimental estimate of the treatment effect from 2013. The partial R^2 of these variables is .033, meaning that they explain 3.3 percent of the residual variation in the combined midterm and final score once the effects of the student characteristics have been partialled out. We cannot reject the null hypothesis that the estimated compressed treatment effect is the same in column (3) and column (5), however.[12]

[11] One student who responded to the survey questions missed the midterm exam for legitimate reasons and therefore is not included in these performance regressions.
[12] To test this hypothesis, we generated 5000 bootstrap samples and with each sample estimated both models. The correlation between the estimates was .927. Using the bootstrap estimates to calculate the covariance between the estimates allows us to calculate a standard error for the difference between the two models, which was 0.0256,

IV. Conclusion

We tested whether class time mattered in an Internet-rich environment with both a randomized and observational design in an effort to simulate a double randomized preference trial (DRPT). A DRPT extends the value of a randomized study by estimating the effect of choosing a treatment among subjects who are same as those who were randomly assigned to treatment. At a minimum, a DRPT offers a test of selection bias by comparing outcomes of the randomized controls to those who chose the control condition in the choice arm.

We found relatively small differences in the effect of class format on academic performance between students in the randomized arm and those who chose their format. We also found little evidence of selection bias in choice of format. One reason may be the nature of the course. Principles of Microeconomics is a required course of all business school majors and almost all non-business majors take it as part of their general education requirements. Students also take the class relatively early in their curriculum because it is a prerequisite for many other classes, accounting for the balance in student characteristics between the fall of 2013 and 2014. The lack of selection in choice of format may be related to the limited flexibility students have in creating a schedule at a commuting college in which many students work. Our survey of students in the observational study revealed few predictors of choice of format and factors such as student GPA—a powerful predictor of academic performance—were unrelated to choice of format. These features of the course and the balance among student characteristics by design and format suggest that we may have met the assumptions of a DRPT. If true, our results suggest that the choice of class format for large required classes is not a major determinant of student performance.

yielding a z-ratio of 1.29. A non-parametric 95 percent confidence interval for the difference between the two models also contained zero.

References

Allen, I. E., & Seaman, J. (2003). *Changing course: Ten years of tracking online education in the United States*. Babson Survey Research Group and Quahog Research Group.

Alpert, W. T., Couch, K. A., & Harmon, O. R. (2015, February). *Online, blended, and classroom teaching of economics principles: A randomized experiment*. Working Paper, University of Connecticut.

Anstine, J., & Skidmore, M. (2005). A Small Sample Study of Traditional and Online Courses with Sample Selection Adjustment. *The Journal of Economic Education*, *36*(2), 107–127.

Awad, M. A., Shapiro, S. H., Lund, J. P., & Feine, J. S. (2000). Determinants of patients' treatment preferences in a clinical trial. *Community Dentistry and Oral Epidemiology*, *28*(2), 119–125.

Bettinger, E., Fox, L., Loeb, S., & Taylor, E. (2014, August). *Changing distributions: How online college classes alter student and professor performance*. Working Paper, Stanford University.

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 Policy Analysis and Management*, *33*(1), 94–111.

Brewin, C. R., & Bradley, C. (1989). Patient preferences and randomised clinical trials. *BMJ*, *299*(6694), 313–315.

Coates, D., Humphreys, B. R., Kane, J., & Vachris, M. A. (2004). "No significant distance" between face-to-face and online instruction: evidence from principles of economics. *Economics of Education Review*, *23*(5), 533–546.

Cohn, E., & Johnson, E. (2006). Class attendance and performance in principles of economics. *Education Economics*, *14*(2), 211–233.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, *27*(4), 724–750.

Devadoss, S., & Foltz, J. (1996). Evaluation of factors influencing student class attendance and performance. *American Journal of Agricultural Economics*, *78*(3), 499–507.

Dobkin, C., Gil, R., & Marion, J. (2010). Skipping class in college and exam performance: Evidence from a regression discontinuity classroom experiment. *Economics of Education Review*, *29*(4), 566–575.

Dunning, T. (2012). *Natural experiments in the social sciences: A design-based approach*. New York: Cambridge University Press.

Durden, G. C., & Ellis, L. V. (1995). The effects of attendance on student learning in principles of economics. *The American Economic Review*, *85*(2), 343–346.

Figlio, D., Rush, Mark, & 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.

Gratton-Lavoie, C., & Stanley, D. (2009). Teaching and learning principles of microeconomics online: An empirical assessment. *The Journal of Economic Education*, *40*(1), 3–25.

Hart, C. M. D., Friedmann, E., & Hill, M. (2015, April 10). *Online course-taking and student outcomes in California community colleges*. Working Paper, University of California, Davis.

Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica*, *66*(5), 1017–1098.

Janevic, M. R., Janz, N. K., Dodge, J. A., Lin, X., Pan, W., Sinco, B. R., & Clark, N. M. (2003). The role of choice in health education intervention trials: a review and case study. *Social Science & Medicine*, *56*(7), 1581–1594.

Johnson, H., & Meijia, M. C. (2014). *Online learning and Student outcomes in California's community colleges*. San Francisco, California: Public Policy Institute of California.

Joyce, T., Crockett, S., Jaeger, D. A., Altindag, O., & O'Connell, S. D. (2015). Does classroom time matter? *Economics of Education Review*, *46*, 64–77.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *The Journal of Economic Perspectives*, *5*(1), 193–206.

Kirby, A., & McElroy, B. (2003). The effect of attendance on grade for first year economics students in University College Cork. *Economic & Social Review*, *34*(3), 311–326.

Lalonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *The American Economic Review*, *76*(4), 604–620.

Little, R. J., Long, Q., & Lin, X. (2008). Comment on: "Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments." *Journal of the American Statistical Association*, *103*(484), 1344–1346.

Long, Q., Little, R. J., & Lin, X. (2008). Causal inference in hybrid intervention trials involving treatment choice. *Journal of the American Statistical Association*, *103*(482), 474–484.

Marburger, D. R. (2001). Absenteeism and undergraduate exam performance. *The Journal of Economic Education*, *32*(2), 99–109.

Marburger, D. R. (2006). Does mandatory attendance improve student performance? *The Journal of Economic Education*, *37*(2), 148–155.

Marcus, S. M., Stuart, E. A., Wang, P., Shadish, W. R., & Steiner, P. M. (2012). Estimating the causal effect of randomization versus treatment preference in a doubly randomized preference trial. *Psychological Methods*, *17*(2), 244–254.

Olitsky, N. H., & Cosgrove, S. B. (2014). The effect of blended courses on student learning: Evidence from introductory economics courses. *International Review of Economics Education*, *15*, 17–31.

Romer, D. (1993). Do students go to class? Should they? *The Journal of Economic Perspectives*, *7*(3), 167–174.

Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in Medicine*, *26*(1), 20–36.

Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, *2*(3), 808–840.

Rücker, G. (1989). A two-stage trial design for testing treatment, self-selection and treatment preference effects. *Statistics in Medicine*, *8*(4), 477–485.

Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association*, *103*(484), 1334–1350.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Stanca, L. (2006). The effects of attendance on academic performance: Panel data evidence for introductory microeconomics. *The Journal of Economic Education*, *37*(3), 251–266.

Wennberg, J. E., Barry, M. J., Fowler, F. J., & Mulley, A. (1993). Outcomes research, PORTs, and health care reform. *Annals of the New York Academy of Sciences*, *703*(1), 52–62.

Xu, D., & Jaggars, S. S. (2011). The effectiveness of distance Education Across Virginia's Community Colleges: Evidence From Introductory College-Level Math and English Courses. *Educational Evaluation and Policy Analysis*, *33*(3), 360–377.

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.

Xu, D., & Jaggars, S. S. (2014). Performance gaps between online and face-to-face courses: Differences across types of students and academic subject areas. *The Journal of Higher Education*, *85*(5), 633–659.

Table 1. Baseline Characteristics of Participants by Academic Year

	2013	2014	Diff.	Norm. Diff.	Log Ratio SD	*N*
Prior Academic Performance						
Baruch GPA	3.03	3.04	-0.01	-0.01	0.03	1102
Transfer GPA	3.31	3.30	0.00	0.01	-0.18	541
SAT Verbal	540.74	556.51	-15.77***	-0.13	0.19	918
SAT Math	604.03	608.73	-4.70	-0.04	0.03	1055
Prior Academic Experience						
Cumulative Credits	44.59	44.88	-0.29	-0.01	-0.01	1333
Part time	0.07	0.08	-0.01	-0.01	-0.03	1333
Underclass	0.76	0.75	0.02	0.02	-0.02	1332
Demographic Characteristics						
Age	20.96	21.08	-0.12	-0.02	-0.20	1333
Female	0.45	0.47	-0.02	-0.02	0.00	1333
White	0.27	0.29	-0.02	-0.03	-0.02	1163
Asian	0.45	0.46	-0.01	-0.02	0.00	1163
Black, Hispanic, Other	0.28	0.25	0.03	0.05	0.04	1163
Native English Speaker	0.53	0.55	-0.03	-0.04	0.00	991
Withdrawal rate	0.10	0.12	-0.02	-0.05	-0.09	1495

Note: This table reports the average background characteristics of students in randomized field experiment (RFE) in Fall 2013 and contrast them with students who enrolled to the same course in the Fall 2014. Sample includes students who completed the course. The column "Diff." shows the difference in means for the indicated variable. Statistical significance means between 2013 and 2014 tested using two sample t-tests assuming unequal variances. Significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$. The column "Norm. Diff" shows the normalized differences, and equals the difference in average covariate values, normalized by the standard deviation of these covariates, i.e. $(\bar{X}_{2003} - \bar{X}_{2004})/\sqrt{s^2_{X,2003} + s^2_{X,2004}}$. The column "Log Ratio SD" shows the logarithm of the ratio of standard deviations and measures of dispersion in the distributions of two covariates. The sample analog of this is calculated as the difference in the logarithms of the two sample standard deviations, i.e. $ln(s_{X,2003}) - ln(s_{X,2004})$. The column "*N*" shows the number of non-missing observations that are used in the comparison.

Table 2. Baseline Characteristics of Participants by Lecture Format and Academic Year

Fall 2013	Compressed	Traditional	Diff.	Norm. Diff.	Log Ratio SD	*N*
Prior Academic Performance						
Baruch GPA	3.01	3.06	-0.05	-0.06	0.04	518
Transfer GPA	3.34	3.28	0.06	0.11	0.06	230
SAT Verbal	544.71	537.12	7.60	0.06	-0.15	511
SAT Math	607.42	600.94	6.48	0.05	-0.10	511
Prior Academic Experience						
Cumulative Credits	45.24	43.96	1.28	0.04	0.12	656
Part time	0.08	0.07	0.02	0.04	0.10	656
Underclass	0.74	0.79	-0.05	-0.09	0.08	655
Demographic Characteristics						
Age	21.23	20.70	0.53**	0.12	0.18	656
Female	0.44	0.46	-0.02	-0.02	0.00	656
White	0.25	0.30	-0.05	-0.08	-0.06	546
Asian	0.46	0.44	0.03	0.04	0.01	546
Black, Hispanic, Other	0.29	0.26	0.03	0.04	0.03	546
Native English Speaker	0.53	0.53	0.00	0.00	0.00	561
			p-value, joint χ^2-test = 0.157			

Fall 2014	Compressed	Traditional	Diff.	Norm. Diff.	Log Ratio SD	*N*
Prior Academic Performance						
Baruch GPA	3.05	3.03	0.02	0.03	0.06	584
Transfer GPA	3.29	3.32	-0.03	-0.05	0.03	311
SAT Verbal	555.20	557.75	-2.55	-0.02	0.03	407
SAT Math	604.92	612.37	-7.45	-0.06	0.00	544
Prior Academic Experience						
Cumulative Credits	44.76	44.99	-0.24	-0.01	0.09	677
Part time	0.07	0.09	-0.01	-0.04	-0.08	677
Underclass	0.73	0.76	-0.02	-0.04	0.03	677
Demographic Characteristics						
Age	21.02	21.14	-0.12	-0.02	0.08	677
Female	0.50	0.44	0.06	0.09	0.01	677
White	0.28	0.30	-0.02	-0.03	-0.02	617
Asian	0.46	0.47	-0.01	-0.01	0.00	617
Black, Hispanic, Other	0.26	0.23	0.03	0.05	0.04	617
Native English Speaker	0.59	0.53	0.06	0.08	-0.01	430
			p-value, joint χ^2-test = 0.615			

Note: This table reports the average background characteristics of students in "compressed" format (lectures once per week) and contrast them with students in "traditional" format (lectures twice per week) for Fall 2013 and Fall 2014 separately. Sample includes students who completed the course during each academic year. The column "Diff." shows the difference in means for the indicated variable. Statistical significance means between 2013 and 2014 tested using two sample t-tests assuming unequal variances. Significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$. The column "Norm. Diff" shows the normalized differences, and equals the difference in average covariate values, normalized by the standard deviation of these covariates, i.e. $(\bar{X}_{2003} - \bar{X}_{2004})/\sqrt{s^2_{X,2003} + s^2_{X,2004}}$. The column "Log Ratio SD" shows the logarithm of the ratio of standard deviations and measures of dispersion in the distributions of two covariates. The sample analog of this is calculated as the difference in the logarithms of the two sample standard deviations, i.e. $ln(s_{X,2003}) - ln(s_{X,2004})$. The column "*N*" shows the number of non-missing observations that are used in the comparison.

Table 3. Student Perfomance

Covariate	Midterm (1)	Midterm (2)	Final (3)	Final (4)	Midterm + Final (5)	Midterm + Final (6)	Course Grade (7)	Course Grade (8)	Withdraw (9)	Withdraw (10)
Compressed (2013)	-0.21***	-0.20***	-0.18**	-0.14**	-0.22***	-0.19***	-0.22***	-0.20***	0.01	0.00
	(0.08)	(0.07)	(0.08)	(0.07)	(0.08)	(0.07)	(0.08)	(0.06)	(0.02)	(0.02)
Compressed (2014)	-0.15*	-0.12*	-0.14*	-0.13*	-0.16**	-0.14**	-0.16**	-0.14**	0.02	0.02
	(0.08)	(0.07)	(0.08)	(0.07)	(0.08)	(0.07)	(0.08)	(0.06)	(0.02)	(0.02)
Diff. (2013 – 2014)	-0.07	-0.08	-0.04	-0.01	-0.06	-0.05	-0.06	-0.07	-0.01	-0.01
	(0.11)	(0.09)	(0.11)	(0.09)	(0.11)	(0.09)	(0.11)	(0.08)	(0.03)	(0.03)
Mon.-Wed.	0.04	-0.05	0.07	-0.01	0.07	-0.03	0.05	-0.05	-0.01	-0.01
	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.02)	(0.02)
Small Classroom	0.26***	0.24***	0.31***	0.30***	0.32***	0.30***	0.30***	0.28***	-0.01	-0.01
	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.02)	(0.02)
Other Covariates		X		X		X		X		X
R^2	0.021	0.342	0.025	0.297	0.029	0.388	0.027	0.426	0.002	0.085
N	1332		1333		1332		1333		1492	

Note: This table reports the differences between student performance in “compressed” format (lectures once a week) and in “traditional” format (lectures twice a week) for the Fall 2013 and Fall 2014 semesters. Coefficients are from the estimation of equation (4) in the text which for convenience we show here. $P_{idf} = \alpha_0 + \alpha_1 C_{if} + \alpha_2 D_{id} + \alpha_3 (C_{id} \times D_{if}) + \sum \beta_k X_{ikdf} + e_{ifd}$. The estimate for the “compressed” lecture format relative to the “traditional” lecture format in 2013 is $\hat{\alpha}_1 + \hat{\alpha}_3$. All outcomes are based on a standardized normal scale with a mean of zero and a standard deviation of 1 within each semester. Estimated with OLS. Heteroskedasticity-consistent standard errors in parentheses. 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. Course Grade includes curved midterm and final grades, penalties for missed classes, and the participation bonus. Significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$.

Table 4. Student Perfomance within Professor/Classroom

	Midterm		Final		Midterm + Final		Course Grade	
Covariate	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Professor A / Small Classroom							
Compressed (2013)	-0.21	-0.12	-0.18	-0.10	-0.21	-0.12	-0.19	-0.12
	(0.15)	(0.13)	(0.15)	(0.13)	(0.15)	(0.12)	(0.15)	(0.12)
Compressed (2014)	-0.25*	-0.11	-0.26*	-0.14	-0.28**	-0.14	-0.31**	-0.15
	(0.14)	(0.14)	(0.14)	(0.13)	(0.14)	(0.13)	(0.14)	(0.12)
Difference	0.04	-0.01	0.09	0.04	0.07	0.02	0.11	0.03
	(0.20)	(0.18)	(0.20)	(0.18)	(0.20)	(0.17)	(0.20)	(0.17)
Other Covariates		X		X		X		X
R^2	0.021	0.341	0.025	0.297	0.029	0.387	0.027	0.426
N	383		383		383		383	
	Professor B / Large Classroom							
Compressed (2013)	-0.19**	-0.27***	-0.14	-0.18**	-0.18**	-0.26***	-0.21**	-0.28***
	(0.09)	(0.07)	(0.09)	(0.08)	(0.09)	(0.07)	(0.09)	(0.07)
Compressed (2014)	-0.08	-0.16**	-0.04	-0.12	-0.06	-0.14**	-0.08	-0.16**
	(0.09)	(0.08)	(0.09)	(0.08)	(0.09)	(0.08)	(0.09)	(0.07)
Diff. (2013-2014)	-0.11	-0.11	-0.10	-0.06	-0.12	-0.10	-0.13	-0.12
	(0.13)	(0.11)	(0.13)	(0.11)	(0.13)	(0.11)	(0.13)	(0.10)
Other Covariates		X		X		X		X
R^2	0.005	0.345	0.003	0.275	0.005	0.378	0.006	0.420
N	949		950		949		950	

Note: This table reports the differences between student performance in "compressed" format (lectures once a week) and in "traditional" format (lectures twice a week) for the Fall 2013 and Fall 2014 semesters within professor/classroom. Capacity of the small classroom is 114 students while the large classroom is 274 students. Coefficients are from the estimation of equation (4) in the text which for convenience we show here. $P_{idf} = \alpha_0 + \alpha_1 C_{if} + \alpha_2 D_{id} + \alpha_3 (C_{id} \times D_{if}) + \sum \beta_k X_{ikdf} + e_{ifd}$. The estimate for the "compressed" lecture format relative to the "traditional" lecture format in 2013 is $\hat{\alpha}_1 + \hat{\alpha}_3$. All outcomes are based on a standardized normal scale with a mean of zero and a standard deviation of 1 within each semester. Estimated with OLS. Heteroskedasticity-consistent standard errors in parentheses. 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. Course Grade includes curved midterm and final grades, penalties for missed classes, and the participation bonus. Estimated with OLS. Heteroskedasticity-consistent standard errors in parentheses. Significance levels are indicated by $* < .1$, $** < .05$, $*** < .01$.

Table 5. Student Self-reported Characteristics by Lecture Format

	Traditional	Compressed	Scale	χ^2-test (p)	N
	(1)	(2)	(3)	(4)	(5)
My learning style is well-suited to a hybrid format	3.29	2.69	[1-5]	0.00***	648
Ever took hybrid/fully online course before	0.32	0.29	[0-1]	0.26	647
Writing focus courses are my strenght	2.94	2.92	[1-5]	0.79	647
I use every available course supplement	2.00	2.05	[1-5]	0.86	648
Economics is not very relevant to my major	3.88	3.94	[1-5]	0.49	646
I typically do not finish my classwork	4.15	4.14	[1-5]	0.50	648
Traditional lectures work well for me	2.30	2.51	[1-5]	0.02**	647
I avoid hard grader professors	2.48	2.47	[1-5]	0.30	648
I need structure to get my class work done	2.02	2.20	[1-5]	0.06*	647
Getting at least A- is a high priority for this class	1.50	1.50	[1-5]	0.19	647
Prefer quantitative courses to writing-focused	2.57	2.54	[1-5]	0.00***	645
Prefer electronic devices to read than paper	3.38	3.23	[1-5]	0.16	647
I am a disciplined person, no need deadlines	2.73	2.75	[1-5]	0.57	646
Commute to campus on weekdays is difficult	3.16	2.95	[1-5]	0.23	646
Economics is a challenging course	2.43	2.45	[1-5]	0.26	646
Interaction with professor and other students helps	1.81	2.12	[1-5]	0.00***	647
Risk preference	6.64	6.78	[0-10]	0.33	648
Commute time to school	2.42	2.52	[1-4]	0.30	647
Paid work during the semester	2.51	2.35	[1-4]	0.15	648

Note: This table reports the differences in pre-class survey responses between the students who chose the "compressed" format (lectures once per week) and the students who chose the"traditional" format (lectures twice per week) during the Fall 2014 semester. Figures in column (1) and (2) are the average score for each question. Column (3) reports the survey question scale. All [1-5] questions used a 5-point Likert scale from 1 to 5 with strongly agree 1, agree 2, neither agree or disagree 3, disagree 4 and strongly disagree 5. The possible answers to the first question is binary and equals one if the answer is "yes". Among the last three questions, the risk preference question has a continuous scale from 1 to 10 and increases in risk-seeking. The commute question has 4 categories: [1] "less than 30 minutes", [2] "between 30 minutes and 60 minutes", [3] "between 60 minutes and 90 minutes" , and [4] "more than 90 minutes". The last question has 4 categories: [1] "No paid work" , [2] "Working less than 15 hours per week", [3] "Working between 15-30 hours per week", and [4] "Working more than 30 hours per week". Column (4) show the p-values from the χ^2-test of independence among responses by format. Column 5 reports the number of non-missing observations for the indicated survey question. Significance levels are indicated by $* < .1$, ** $< .05$, *** $< .01$..

Table 6. Lecture Format Choice and Student Performance in Fall 2014

Covariate	"Compressed" Coef.	Std. Err.	Midterm+Final Coef.	Std. Err.	Midterm+Final Coef.	Std. Err.
	(1)	(2)	(3)	(4)	(5)	(6)
Compressed (2014)			-0.158	0.064*	-0.191	0.068**
Baruch GPA	0.007	0.037	0.657	0.067***	0.649	0.067***
Transfer GPA	0.074	0.063	0.236	0.104*	0.204	0.105
SAT Verbal	-0.000	0.000	0.001	0.001	0.001	0.001
SAT Math	0.000	0.000	0.004	0.001***	0.004	0.001***
Cumulative Credits	0.002	0.002	-0.000	0.003	0.000	0.003
Part time	0.025	0.075	-0.027	0.130	-0.024	0.131
Underclass	0.086	0.086	0.063	0.154	0.062	0.153
Age	0.002	0.007	0.001	0.012	-0.001	0.012
Female	-0.062	0.039	-0.335	0.067***	-0.315	0.067***
Asian	0.016	0.049	-0.074	0.081	-0.063	0.082
Black, Hispanic, Other	0.017	0.057	-0.090	0.097	-0.086	0.100
Native English Speaker	-0.052	0.049	-0.085	0.084	-0.072	0.083
My learning style is well-suited to a hybrid format						
Neutral	-0.134	0.047**			0.055	0.077
Agreed	-0.304	0.049***			-0.021	0.089
Traditional lectures work for me						
Neutral	0.022	0.044			0.069	0.069
Agreed	0.076	0.061			0.008	0.107
I need structure to get my class work done						
Neutral	0.003	0.056			0.066	0.095
Agreed	0.107	0.068			0.173	0.114
Prefer quantitative courses to writing-focused						
Neutral	0.029	0.043			-0.249	0.073***
Agreed	-0.108	0.057			0.280	0.092**
Interaction with professor and other students helps						
Neutral	0.146	0.056**			0.129	0.107
Agreed	0.132	0.080			-0.086	0.121
N	677		676		676	
R^2	0.120		0.333		0.355	

Note: The dependent variable in column (1) is a dichotomous indicator that is 1 if the student chose the "compressed" format and 0 if she chose the "traditional format". The dependent variable in columns (3)-(6) is the score on the combined midterm and final standardized with mean zero and standard deviation of one. Estimated with OLS. Heteroskedasticity-consistent standard errors are in the adjacent column with the significance levels, indicated by $* < .1$, $** < .05$, $*** < .01$.