Requiring Versus Recommending Preparation Before Class: Does It Matter?

By: Martin S. Andersen, Dora Gicheva, and Jeffrey Sarbaum

This is the peer reviewed version of the following article:

Andersen, M. S., Gicheva, D., & Sarbaum, J. (2018). Requiring versus recommending preparation before class: Does it matter? *Southern Economic Journal, 85*(2), 616-631.

Which has been published in final form at http://dx.doi.org/10.1002/soej.12281

Abstract:

Asking students to come to class prepared is quite common in undergraduate and graduate education. We use a quasiexperimental design to assess whether requiring undergraduate students in an introductory course to review prior to lecture the material that will be taught in class enhances their understanding of key concepts. We find that requiring rather than recommending preparation before class increases exam scores by about a quarter of a standard deviation, or roughly a third of a letter grade, for students in the second and third quartiles of the ability distribution but has little impact on very high- or low-ability students. We also estimate local average treatment effects, from which we draw a similar conclusion: reviewing the material before lecture benefits students in the middle of the ability distribution but has essentially no impact on the top and bottom quartiles.

Keywords: class preparation | classroom instruction

Article:

1. Introduction

Many instructors expect students to respond to requests asking them to come to class having completed preparatory work. However, behavior toward such requests and the role that completing preparatory work before class has on student learning has been largely overlooked in the literature. In this article, we use a quasiexperimental design to assess whether requiring students in a Principles of Microeconomics course to review prior to lecture the material that will be taught in class enhances their understanding of key concepts compared to requiring students to review the same material at some point before the semester ends. Further, we investigate whether preparation before class has heterogeneous impacts depending on students' ability levels.

The research questions we address cannot be answered reliably by comparing the performance of students who choose to review the material prior to class to the success rates of students who do similar amounts of course work but after lecture because of the potential for unobserved confounding factors. Students with high levels of unobserved ability, work ethic, or motivation may be more likely to complete reading assignments prior to lecture, and at the same time are likely to perform better in their classes, generating a spurious positive correlation between preparation and test scores. Similarly, students who are more future-oriented may be more likely to complete assignments early, and this trait has also been linked to improved academic performance in college (Kirby, Winston, and Santiesteban 2005). Conversely, students who struggle with the content and are thus likely to have lower test scores may also be more likely to spend time reviewing the material prior to lecture. The latter would result in negative correlation between preparation before class and performance.

To avoid unobserved heterogeneity and simultaneity bias, we implemented an experiment in which we assigned sections of a Principles of Microeconomics course taught in the Spring 2017 semester at the University of North Carolina at Greensboro (UNCG) to one of two conditions. Roughly half of the students were required to complete assignments prior to lecture, while instructors recommended but did not require such preparation for the rest of the students in the course. The second group still had to complete the assignments as part of their course grade, but had until the end of the semester to do so. Each assignment not completed by the assigned due date resulted in a reduction of slightly over one percentage point of the final course grade; no partial credit was given to either group for late submissions.

The condition to which their section was assigned played a major role in the timing of assignment completion for most students in the course. Requiring students to prepare prior to class increased the probability that a student prepared prior to class by 60–70 percentage points. We estimate both the intent-to-treat effect of assigning work prior to lecture and the local average treatment effect of preparing before class on student performance. We find that requiring rather than recommending preparation before class increases exam scores by about a quarter of a standard deviation, or roughly a third of a letter grade, for students in the second and third quartiles of the ability distribution but has little impact on very high- or low-ability students. The local average treatment effects we estimate are somewhat larger but tell the same story: reviewing instructor-developed reading materials and videos and answering related multiple-choice questions before lecture benefits students in the middle of the ability distribution but has essentially no impact on the top and bottom quartiles.

Using experimental designs, several recent studies find evidence in support of assigning homework in Principles of Microeconomics courses (Grove and Wasserman 2006; Emerson and Mencken 2011; Grodner and Rupp 2013). At the same time, Harter and Harter (2004) show that simply making additional online quizzes available without monitoring completion does not improve student performance in an introductory economics course. Others have found support for using "pre-lecture" materials to create classroom time for critical thinking exercises (Balaban, Gilleskie, and Tran 2016; Caviglia-Harris 2016; Swoboda and Feiler 2016). Similar research in the STEM disciplines has found that preclass preparation is positively related to exam performance. For example, Seery and Donnely (2012) compare exam scores in an introductory chemistry class over the course of several semesters and find that making prelecture resources available and assigning quizzes before class bridges the performance gap between students with and without high school chemistry exposure. Moravec et al. (2010) compare student performance in a large introductory biology class in two consecutive semesters to show that moving some of the material from lecture to preclass worksheets and videos leads to exam score improvements. At the same time, active-learning exercises structured as frequent low-stake assessments are found to lower failure rates and increase exam scores in introductory biology and psychology courses (Freeman et al. 2007; Freeman, Haak, and Wenderoth 2011; Haak et al. 2011; Pennebaker, Gosling, and Ferrell).

While the use of instructor-developed problem sets is common in introductory economics courses (Watts and Becker 2008; Watts and Schaur 2011), and prior research provides compelling evidence for homework completed after lecture and for preclass assignments, it does not directly address the question of whether requiring some of the work to be completed prior to lecture impacts the education production function. All students in our study were graded based on the same set of assignments, but the timing of the due dates differed so that in roughly half of the course sections, a set of assignments had to be completed prior to lecture. The design of the experiment we conducted allows us to address the novel question of how restructuring due dates while keeping the overall workload fixed impacts students in an introductory economics course.

Our study uses quasiexperimental methodology that avoids several limitations of other studies of the factors that go into the education production function. The course we study had a fairly large number of sections (15) all taught in the same semester by a total of 5 instructors. Every instructor taught at least one section in each treatment condition. This allows us to ensure that semester and instructor effects do not drive the results. Even though students were not randomly assigned to sections, they were unaware of the experiment at the time of registration, which combined with variations in the treatment condition and instructor within the time of day when classes met is likely to eliminate any potential selection bias.

The rest of the article is organized as follows. Section 2 provides more details about the setting for our study. We describe the data set we use for the analysis in section 3, and the empirical models we estimate in section 4. We present the estimation results in section 5, and section 6 concludes.

2. Course Setting and Estimation Sample

The Principles of Microeconomics course is taught as a hybrid part online part face-to-face class, in which students meet with an instructor once a week for 75 minutes. In addition, there are online materials available including text and videos developed by one of the faculty instructors for the course.[1] Students are required to review the online materials each week and to take an online “check-in quiz.” While check-in quizzes are typically due the night before the face-to-face lecture, for the purposes of the experiment analyzed in this study, we changed the requirements for some sections in the Spring 2017 semester so the quizzes were due at the end of the semester. Instructors strongly and repeatedly recommended during class that students complete the assignments before lecture even though they were not due until after the final exam, and the course calendar in Canvas, UNCG's Learning Management System, had “recommended be done by date” reminders that mirrored how the “due by date” reminders were displayed in the other sections. Answers to these check-in assignments are readily available as part of the online materials, although not in one place but scattered around the materials to encourage students to actually go through the reading and videos. Most students who take the quizzes obtain a score of 100% or close.

While the nine check-in quizzes account for 10% of the course grade, assessment is mainly based on two midterm exams worth a total of 40% of the final course grade and a final exam accounting for 30% of the grade.[2] All assignments, including exams, are identical across all sections of the class. The first midterm covers the material on check-in quizzes 1 through 4; the second midterm covers the material from quizzes 4 through 7, and the final is comprehensive. The format of all exam questions is multiple choice.

Capacity in each course section is capped at 32 students and sections were full at the start of the Spring 2017 semester, meaning that there were 256 students who registered for the eight sections in which assignments were due prior to lecture (referred to as "before" sections in the rest of the paper). There were 224 students who started in the 7 sections in which instructors recommended that students complete the assignments prior to class but with actual due dates were 24 hours after the final exam (referred to as "end" sections from here on). Students had no information at the time of registration about the differences in the structure of assignment due dates or the fact that instructors were conducting any sort of experiment. There were five instructors, two full-time faculty and three Ph.D. students, each of whom taught at least one "before" and one "end" section. The Appendix shows more details about the assignment of instructors to sections and about the times of day when each section met.

One distinctive institutional feature important for our study is the course withdrawal policy at UNCG. Beginning with the 2014–2015 academic year, undergraduate students were restricted to withdrawing from no more than 16 credit hours during their undergraduate career, except for the first week of the semester when students could freely add and drop courses. Students who have not dropped more than 16 semester hours and drop a course in the first eight weeks of the semester receive a grade of WX (withdrawn) on their transcript that does not count toward their grade point average; later withdrawals or withdrawals in excess of 16 semester hours result in a grade of WF (withdrawn failing) on the student's transcript, which counts as a grade of "F" in the GPA calculation. As a result, withdrawal rates at UNCG have been very low, and while previous studies of economics students' performance such as Becker and Powers (2001) have emphasized the importance of taking attrition into account, this is less of a concern it the current setting.

Of the 480 students who started the class, 463 remained enrolled past the first-week drop-add period; 459 took the first exam, 5 of whom did not take the second midterm. There are 449 students who took the second midterm and also took the final exam,[3] which puts overall attrition at 6.5%; the difference in sample attrition rates between the two types of sections is not statistically different from zero. For the rest of the analysis, we assume that attrition is uncorrelated with the type of section students were enrolled in, but even if this assumption is violated, attrition is low and should not affect the results. Our final estimation sample consists of the 237 students in "before" sections and the 212 students in "end" sections who took all three exams.[4] We conduct most of the analysis at the student-exam level, for which we have 1347 observations.

3. Variables Used in the Analysis

We combine data collected by instructors containing course outcomes such as assignment completion rates, absences, and test scores for all students who completed the class with information about student demographics, entrance exam scores, and grade point average provided by the University's Office of Institutional Research.

A main explanatory variable of interest is the fraction of check-in quizzes that each student completed prior to lecture. Since students were provided with the answers to check-in quizzes, we use only an indicator of assignment completion rather than the actual scores, for which there is very little variation. As Table 1 demonstrates, completion rates varied greatly between section types, with about 90% of students in "before" sections completing the quizzes prior to class, and between 17 and 29% of students in "end" sections doing so.5 Assignment completion is also highly correlated within student, especially in the "end" sections; the correlation is somewhat lower in "before" sections, where few students missed assignments systematically.

Table 1. Check-in Assignment Completion Rates

	Share Completed Before Lecture		
Assignment Number	"Before" Sections	"End" Sections	*t*-statistics
1	0.91	0.28	18.02
2	0.92	0.25	19.54
3	0.94	0.23	22.53
4	0.89	0.24	18.24
5	0.89	0.22	19.35
6	0.89	0.17	22.13
7	0.88	0.20	19.78
8	0.89	0.18	21.57
9	0.87	0.29	15.50

N = 237 in "before" sections (required to complete assignments before class) and N 5 212 in "end" sections (required to complete assignments by the end of the semester). The t-statistics are from tests of equal means between the two groups.

To describe in more detail the within- and across-student variation in the probability of completing an assignment before class, we estimate separately by section type a random effect model for student i completing assignment c:

$$Complete_{ic} = \mu + u_i + e_{ic},$$

where $1 \leq c \leq 9$. We use the results from this model to calculate intraclass correlation coefficients

$$\hat{\rho} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2},$$

where $\hat{\sigma}\frac{2}{u}$ and $\hat{\sigma}\frac{2}{e}$ are, respectively, the estimated variances of u_i and e_{ic}. The estimated inctraclass correlation for the "end" group is 0.52, while for the "before" group it is 0.21 and the difference in the intraclass correlation coefficients is statistically significant at conventional significance levels. Exam scores are also highly correlated for each student; using a similar method, we estimate the intraclass correlations to be 0.65 for the "before" group and 0.71 for the "end" group, but in this case we observe no statistically significant difference in the intraclass correlation coefficients of the two groups.

We use SAT and ACT scores, as well as high school GPA, to construct a measure of ability as follows. We first standardize each of the three measures to have a mean of 0 and standard deviation of 1 for all students who have a nonmissing value for that variable. Next, we set the ability measure, which we refer to as pretest score in the rest of the paper, to equal the standardized SAT score for all students for whom it is available (67.2% of the sample). For students who do not have an SAT score, we use the standardized ACT score when available (17.2% of the sample). Finally, we set the pretest measure equal to the standardized high school GPA if it is available but SAT and ACT scores are not (3.2% of the sample). We impute pretest scores for the remainder of the sample using multiple regression imputation on age and indicators for gender, Hispanic ethnicity, race, in-state residency, whether the student lives on campus, citizenship status, and major field of study, as well as UNCG grade point average at the start of the semester for students who had completed prior course work at the University.

Table 2 shows descriptive statistics for student demographics and achievement measures separately by section type. Column 3 shows t-statistics from the test of equal means for the two groups. It can be seen from Table 2 that for all but one of the variables the null hypothesis that the means are equal cannot be rejected at the 10% significance level; we interpret this as evidence that even though student assignment was not entirely random, there is no systematic selection on observables. While selection on unobservables cannot be ruled out, it is much less likely when the samples are matched closely on observed characteristics.

Almost half of students in the sample have first-year status, and about the same fraction live in university housing. Close to half of students are non-white and non-Asian, which reflects the composition of the UNCG student body, which is diverse relative to most four-year universities in the United States Since the introductory microeconomics course is required for all majors in the business school, and Economics is a relatively small major, there are very few Economics majors in the class. Only 32 students in the sample do not have a declared major. Other concentrations grouped in the excluded category include those in the humanities, other social sciences, and education. Table 2 also shows that students in "before" and "end" sections have similar scores on the pretest measure we construct. The students in "before" sections have slightly higher high school and college GPA, although the differences are not statistically significant, while students in the "end" sections have very small advantage on SAT and ACT scores.

Table 3 shows descriptive statistics for students' performance in the course; we split the sample again by the type of section students were enrolled in, but also examine the outcomes by quartile of the pretest distribution to match more closely the structure of the main empirical models. Attendance rates are high for all groups; the maximum number of lectures a student can get credit for is 11, and the mean for each group is around 10.6 A number of previous studies have linked attendance to student performance in introductory economics courses (e.g., Marburger 2001; Stanca 2006). Students completed most homework assignments, and completion rates did not vary by ability. We include homework completion as a proxy for the effort students put into the class but do not control for homework scores because they may be affected by preparation prior to class, our variable of main interest. The difference in attendance and homework completion rates between "before" and "end" sections is not statistically significant for any of the ability groups. Both are slightly higher in "before" sections for the second quartile, but this is reversed for the third quartile. Among students with the lowest pretest scores, the difference in attendance is negligible, and homework completion is slightly higher in "before" sections.

Table 2. Descriptive Statistics

	(1) "Before Sections	(2) "End" Sections	(3) *t*-statistic
Female	0.44	0.54	-2.11*
Age	19.78	19.92	-0.46
	(2.626)	(3.854)	
Freshman	0.49	0.49	-0.02
Sophomore	0.31	0.3	0.24
Junior	0.15	0.17	-0.64
Lives on-campus	0.53	0.52	0.17
Hispanic	0.07	0.08	-0.34
Black	0.3	0.33	-0.6
Asian	0.1	0.1	-0.07
Other race, non-white	0.1	0.09	0.42
In-state resident	0.91	0.9	0.38
Declared major:			
Accounting	0.11	0.08	0.74
Science	0.03	0.04	-1.01
Business	0.35	0.33	0.46
Economics	0.02	0.03	-1.1
Finance	0.08	0.05	1.2
Information Systems	0.07	0.11	-1.36
Marketing	0.08	0.09	-0.36
U.S. citizen	0.93	0.91	1.04
Standard pretest score (including imputed)	0.04	0.05	-0.14
	(0.945)	(0.956)	
Missing pretest score	0.14	0.12	0.65
UNCG hours at start of semester	24.73	25.03	-0.15
	(21.93)	(19.31)	
GPA at start of semester	3.06	2.99	1.09
	(0.659)	(0.681)	
SAT (z-score)	0	0.02	-0.15
	(0.986)	(1.013)	
ACT (z-score)	-0.01	0.05	-0.39
	(0.961)	(1.035)	
High school GPA	3.79	3.73	1.13
	(0.446)	(0.487)	

*p < 0.1. Standard errors in parentheses. The t-statistics are from tests of equal means between the two groups. N = 449 students.

Scores on the three exams differ very little between the two types of sections for students in the lowest quartile of the ability distribution. In the quartile above that, scores are consistently higher for the "before" sections, but only one of the differences is statistically significant. In the second-highest quartile, students in "before" sections scored higher on the first midterm and the final exam compared to their peers in "end" sections. In the top quartile, there is little difference in the scores on the last two exams, and the "before" group appears to do a little better on the first midterm. The evidence in Table 3 points to an interaction between student ability and the impact of requiring preparation before class, which we investigate further. Most of the differences in Table 3 are not statistically significant, so it is important to add the student controls from Table 2 in order to estimate the impact of the policy with more precision.

We analyze the data as a balanced panel with three entries per student, one for each exam. We calculate and use z-scores for each test rather than the raw score because scores on the final exam are lower and vary less than midterm grades. The fraction of check-in quizzes that a student completes is a main covariate of interest, and there are several ways to incorporate this measure into the longitudinal data set. Our approach is to calculate the fraction of completed quizzes covering the new material tested with each exam. In other words, the check-in score for period 1 equals the average completion rate of the first 4 check-in quizzes; the score in period 2 equals the average of quizzes 5 through 7; and the period 3 score is based on quizzes 8 and 9. Since completion rates do not vary much by student, as discussed earlier in this section, redefining this variable does not impact the results much. In the following section, we discuss the empirical models that we estimate, and section 5 presents the results.

Table 3. Course Performance by Student Ability

Pretest Percentile		"Before" Sections	"End" Sections	*t*-statistic
<25th	Number classes attended	10.22	10.25	-0.16
(N = 113)		(1.027)	(0.985)	
	Number missed HW	1.45	1.2	0.62
		(2.036)	(2.189)	
	Midterm 1	53.53	54.7	-0.38
		(17.42)	(15.23)	
	Midterm 2	58.32	58.82	-0.18
		(14.93)	(14.56)	
	Final exam	53.24	52.99	0.1
		(12.62)	(13.04)	
25th to 50th	Number classes attended	9.89	10.13	-0.93
(N = 124)		(1.732)	(0.972)	
	Number missed HW	1.51	1.11	1.02
		(2.412)	(1.829)	
	Midterm 1	65.07	58.94	2.16*
		(16.35)	(14.83)	
	Midterm 2	64.36	62.08	0.89
		(14.82)	(12.99)	
	Final exam	59.77	55.47	1.82
		(12.85)	(13.31)	
50th to 75th	Number classes attended	10.14	9.86	1.12

(N = 108)		(0.990)	(1.549)	
	Number missed HW	1.14	1.31	-0.46
		(1.894)	(2.035)	
	Midterm 1	67.94	63.63	1.43
		(15.23)	(15.98)	
	Midterm 2	65.44	66.32	-0.29
		(16.30)	(15.87)	
	Final exam	64.56	61.41	1.15
		(13.46)	(15.08)	
>75th	Number classes attended	10.15	10.19	-0.16
		(1.349)	(1.085)	
	Number missed HW	1.35	1.46	-0.26
		(2.266)	(2.271)	
	Midterm 1	76.59	74.18	0.89
		(13.31)	(14.21)	
	Midterm 2	73.37	74.47	-0.4
		(15.00)	(12.77)	
	Final exam	72.02	71.91	0.04
		(13.55)	(12.69)	

*$p < 0.1$. Standard deviations in parentheses. The t-statistics are from tests of equal means between the two groups.

4. Empirical Specifications

There are two related questions, both of which we believe are important, that can be addressed with the data. The first is a matter of intent-to-treat (ITT): Does requiring preparation before class improve student performance? The answer to this question is of interest to instructors considering implementing a similar policy. We also estimate a local average treatment effect (LATE), which addresses a different question: Does reviewing prior to lecture the material that will be taught in class enhance learning for students who only complete the assignments if they are required rather than recommended? Because completion rates are very low in the "end" sections and close to 90% in the "before" sections, the LATE estimates will be an approximation of an average treatment effect. The answer to the latter question will tell us more about the mechanisms behind student learning.

In our main analyses, we account for the possibility that the benefits of requiring preparation before class vary with students' ability levels. We would expect that the highest-ability students are more likely to have developed study strategies that fit their personal approaches to learning and may be relatively less likely to benefit from structured preparation plans such as the one we discuss. Low-achieving students may not benefit from required preparation before class if the level of the material is too high or if they are negligent when working on the assignments. Given that answers to the quiz questions were available to students even if they did not read the material, we believe the latter to be a driving factor for the course we analyze.

To identify the ITT, we estimate the following model for the standardized score of student i on exam j under instructor t:

$$Score_{ijt} = X_i'\beta + \sum_{k=1}^{4} a_k Pretest\_k_i + \sum_{k=1}^{4} \gamma_k (Before_i \times Pretest\_k_i) \qquad (1)$$

$$+\theta_t + \varepsilon_{ijt},$$

where $Pretest\_k_i$ is an indicator for student i being in pretest group k, with group 1 being the lowest-achieving one and group 4—the highest-achieving. The variable $Before_i$ is an indicator for being in a "before" section. The vector of student-level controls in x_i includes age and indicators for gender, Hispanic ethnicity, race, in-state residency, whether the student lives on campus, citizenship status, and major field of study. We also include the imputed pretest scores and an indicator for the imputed observations. The number of lectures the student attended and the homework assignments she completed are in the list of covariates as well; in principle these variables may have been affected by the treatment, but the evidence in Table 3 suggests that this is not the case and models that exclude these variables produce similar estimates to those from our main specification. Where appropriate, we include an indicator equal to 1 if the section is taught by a graduate student. This specification is similar to the ones used in other studies of the effectiveness of various teaching methodologies, such as Emerson and Taylor (2004) and Grove and Wasserman (2006).

We estimate Equation 1 θ_t either as a random or a fixed instructor effect. While students are not assigned randomly to instructors, there is a degree of randomness more so than in other contexts analyzed in the literature. As we point out in section 3, most students taking this Principles of Microeconomics class do not major in economics, and in addition the sample median for the number of credits taken at UNCG prior to the start of the current semester is 15, corresponding to 5 standard courses. Hence, we expect many students to be unfamiliar with the instructors' names when registering. Furthermore, demand for seats in almost all sections typically exceeds the fixed capacity, which means that many students are forced to register for whichever section has seats available. Finally, the Appendix shows that there is variation in the times of day when each instructor teaches, so students who were targeting a specific starting time could end up in different instructors' sections. Thus, we believe that the random effects assumption is reasonable in this case; other quasiexperimental studies such as Emerson and Taylor (2004) make a similar assumption. We also verify that our results are robust to estimating θ_t as a fixed effect. In this model, we treat the three observations for each student as independent. As another robustness check, we estimate the model in Equation 1 only for the scores on the cumulative final exam.

We are also interested in the following regression:

$$Score_{ijt} = X_i'\alpha + \sum_{k=1}^{4} b_k Pretest\_k_i + \sum_{k=1}^{4} \delta_k \left(PctComplete_{ij} \times Pretest\_k_i\right) + \rho_t + \epsilon_{ijt}, \tag{2}$$

where *PetComplete*$_{ij}$ is the fraction of check-in quizzes that student i completed prior to exam j. The coefficient δ_k measures the average treatment effect for group k of completing the assignment prior to class rather than later in the semester. Earlier in this article we discuss the reasons for which the *PetComplete*$_{ij}$ variable is likely to be correlated with the error term, so we estimate two-stage least squares model in which we use being in a "before" section to instrument for the fraction of completed assignments. Thus, we estimate the LATE for each pretest group. Given the limited within-student variation in the outcome variable and the main explanatory variable of interest, we do not include student fixed effects in the model in Equation 2, but verify the robustness of the results to using only the final exam scores as an outcome variable. We estimate a generalized two-stage least squares model with instructor random effects but verify the robustness of the results to including instructor-level fixed effects instead. We next present the estimation results for the ITT and LATE models.

5. Results

Table 4 shows results for the ITT analysis following the empirical specification in Equation 1. Column 1 shows the average effect of assigning work before class on all students, while the specifications in columns 2–4 explore the treatment effect heterogeneity by quartile of the pretest score distribution. The specifications in columns 2 and 4 include instructor random effects; instructor-level fixed effects are included in the model in column 3. The last column restricts the estimation to final exam scores.

The estimated coefficient on the "before" section indicator in column 1 suggests that when students are required to prepare prior to lecture, average test scores increase by 0.14 standard deviations, or about 2 percentage points. This estimate is highly significant. Looking more closely at the estimated effects by ability, the estimates in column 2 show essentially no impact on students in the lowest and highest quartiles of the distribution of achievement test scores, but an increase of 0.26 standard deviations for students in the second quartile and 0.29 standard deviations for those in the third quartile of the ability distribution. Both estimates are significant at the 1% level in the model with instructor random effects and suggest that conditional on student characteristics and measures of effort, requiring preparation before class increases test scores by about 3–4 percentage points for students whose precollege ability falls between the 25th and 75th percentiles of the distribution for the course.

Including instructor-level fixed effects in the model attenuates the estimates, especially for the second quartile of pretest scores where we obtain a point estimate of 0.16 that is significant at the 10% level. The point estimate for the third quartile decreases from 0.29 to 0.25

and remains significant at the 1% level. Focusing on final exam scores yields point estimates of about 0.3 for the middle two groups and estimates closer to 0 for the lowest- and highest-ability groups; not surprisingly, the standard errors in this model are higher given the smaller number of observations.

The other estimates shown in Table 4 are not surprising and suggest that effort, as measured by attendance and homework completion, is positively correlated with performance, and so is the measure of ability we use. In addition, test scores were higher in sections taught by a full-time faculty member rather than a graduate student, even though all instructors shared identical course materials.

Table 4 Intent-to-Treat Estimation Results

Model:	(1) Instructor RE	(2) Instructor RE	(3) Instructor FE	(4) Final Exam Only and Instructor RE
"Before" section	0.137***			
	(0.046)			
"Before section x Pretest group 1		-0.031	-0.053	0.015
		(0.092)	(0.092)	(0.157)
"Before" section x Pretest group 2		0.286***	0.247***	0.319**
		(0.091)	(0.091)	(0.156)
"Before" section x Pretest group 3		0.259***	0.162*	0.303*
		(0.095)	(0.096)	(0.161)
"Before" section x Pretest group 4		0.026	-0.008	-0.00001
		(0.096)	(0.096)	(0.164)
Pretest group 1 (<25^{th} percentile)		-0.294*	-0.312*	-0.496*
		(0.165)	(0.164)	(0.280)
Pretest group 2 (25^{th} to 50^{th} percentile)		-0.053	-0.061	-0.260
		(0.131)	(0.130)	(0.223)
Pretest group 3 (50^{th} to 75^{th} percentile)		-0.126	-0.177	-0.161
		(0.111)	(0.111)	(0.189)
Number classes attended	0.104***	0.100***	0.129***	0.135***
	(0.020)	(0.020)	(0.020)	(0.033)
Number missed HW	-0.047***	-0.049***	-0.043***	-0.049**
	(0.011)	(0.011)	(0.011)	(0.019)
Section taught by Ph.D. student	-0.182***	-0.188***		-0.167*
	(0.050)	(0.050)		(0.086)
Standardized pretest score	0.400***	0.315***	0.308***	0.278***
	(0.027)	(0.057)	(0.057)	(0.097)
N	1347	1347	1347	449

*p < 0.1, **p < 0.05, ***p < 0.01. All models include age and indicators for gender, Hispanic ethnicity, race, in-state residency, whether the student lives on campus, citizenship status, major field of study, and imputed pretest score.

Before discussing our LATE estimation results, we show first-stage estimation results in Table 5 that describe the relationship between section type and the share of completed assignments, conditional on student characteristics. In order to show an F-statistic for each of the four pretest groups, the results presented in Table 5 are from models estimated separately for each quartile of the ability distribution; the results for the full sample are similar. In all four cases, the predictive power of the instrument is very strong, consistent with the descriptive statistics in Table 1. The relationship between section type and assignment completion is weakest in the highest-ability group, where the estimated first-stage coefficient on the indicator for "before" section is 0.58, and the F-statistic of the instrument is 275. The coefficient estimates are similar for the other 3 groups, ranging between 0.68 and 0.70. The estimates are similar to the raw difference between the two types of sections shown in Table 1 and further suggest that the requiring preparation before lecture is much more effective in incentivizing students to read the material compared to recommending preparation before class but setting a due date later in the semester.

Table 5. First-Stage Estimation Results

Pretest Percentile	(1) <25^{th}	(2) 25^{th} to 50^{th}	(3) 50^{th} to 75^{th}	(4) >75^{th}
"Before" section	0.675***	0.698***	0.683***	0.581***
	(0.034)	(0.034)	(0.033)	(0.035)
F-statistic of instrument	397.54	426.55	440.42	274.99
N	339	372	324	312

***$p < 0.01$.

Table 6 presents our estimates of the local average treatment effect of preparation based on Equation 2 and section type as an instrument for the fraction of completed assignments. The first two columns present generalized least squares (GLS) estimates of this relationship with instructor random effects. The remaining four columns present estimates from alternative two-stage instrumental variable (IV) models. Column 1 demonstrates that completing all the check-in quizzes prior to class is associated with an almost 0.2 standard deviation higher score on the subsequent exam relative to not completing any assignments prior to lecture. Column 2 breaks down the GLS relationship by quartiles of pretest ability and indicates that preparing for class is associated with lower test scores for students in the lowest quartile of pretest ability. This negative relationship may be due to reverse causality: some students who were aware that they did not have a good understanding of the material and were performing poorly in the class may have completed the assignments prior to lecture in an attempt to catch up and improve their exam scores. The coefficients on the second and third quartile interactions indicate that preparation is associated with higher exam scores for individuals in these quartiles of pretest ability. This could reflect a productivity effect of preparation or omitted variable bias if students who prepared for class were also students who were better on unobservable dimensions correlated with academic achievement. In order to distinguish between the two explanations, we use the type of section students were enrolled in as an instrument for the fraction of assignments they completed prior to lecture.

The instrumental variable results that include an instructor random effect (columns 3 and 4) demonstrate that, on average, preparing for class increases exam scores (column 3). Allowing for heterogenous effects by pretest ability yields more nuanced inferences. For students in the lowest quartile, there appears to be no benefit from preparing for class. Combined with the GLS results, this is consistent with low-ability students, or students who are not strong test-takers, doing slightly more work to prepare for class when recognizing that they have a poor understanding of the material, but at the same time not benefitting much from instructor-developed online reading and video materials. In the second and third quartiles, we do find evidence that preparing for class increases subsequent exam scores. Furthermore, comparing the GLS and IV results (columns 2 and 4) indicates that there are two different explanations for the positive GLS coefficients. In the second-lowest quartile, the GLS estimate is biased downward, similarly to the GLS estimate for the lowest-ability group. This is again suggestive evidence that poorer performing students in this group were more likely to prepare for class. In the second-highest pretest quartile, it was students who perform better who were more likely to prepare for class leading to an upward omitted variable bias in the GLS coefficient.

Table 6. Local Average Treatment Effects

	(1) GLS	(2) GLS	(3) IV	(4) IV	(5) IV	(6) IV
% completed check-ins	0.177***		0.202***			
			(0.068)			
Completed check-ins		-0.168*		-0.049	-0.079	0.020
X Pretest group 1		(0.102)		(0.136)	(0.136)	(0.231)
Completed check-ins		0.217**		0.398***	0.347***	0.437**
X Pretest group 2		(0.099)		(0.129)	(0.129)	(0.216)
Completed check-ins		0.544***		0.361***	0.228*	0.419*
X Pretest group 3		(0.105)		(0.133)	(0.135)	(0.223)
Completed check-ins		0.084		0.037	-0.021	-0.009
X Pretest group 4		(0.110)		(0.170)	(0.169)	(0.295)
Pretest group 1 (<25th		-0.118		-0.216	-0.255	-0.526
percentile)		(0.174)		(0.199)	(0.198)	(0.341)
Pretest group 2 (25th to		-0.251*		-0.384**	-0.393**	-0.672**
50th percentile)		(0.138)		(0.164)	(0.163)	(0.282)
Pretest group 3 (50th to		-0.501***		-0.418***	-0.396**	-0.550**
75th percentile)		(0.125)		(0.156)	(0.155)	(0.268)
Number classes	0.101***	0.095***	0.100***	0.097***	0.127***	0.131***
Attended	(0.020)	(0.020)	(0.020)	(0.020)	(0.021)	(0.033)
Number missed HW	-0.042***	-0.044***	-0.041***	-0.045***	-0.041***	-0.044**
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.020)
Section taught by grad.	-0.165***	-0.172***	-0.164***	-0.173***		-0.150*
Student	(0.050)	(0.050)	(0.050)	(0.050)		(0.085)
Standardized pretest	0.401***	0.312***	0.402***	0.316***	0.309***	0.277***
score	(0.027)	(0.057)	(0.027)	(0.057)	(0.057)	(0.097)
N	1347	1347	1347	1347	1347	449

*p < 0.1, **p < 0.05, ***p < 0.01. The IV models use section type as an instrument for the fraction of completed check-in assignments. All models include age and indicators for gender, Hispanic ethnicity, race, in-state residency, whether the student lives on campus, citizenship status, major field of study, and imputed pretest score. The specifications include instructor random effects, with the exception of the one in column 5, which includes instructor-level fixed effects. The estimation in column 6 is limited to final exam scores.

There is no evidence that preparation improves performance for people in the highest quartile of the ability distribution, despite the fact that being assigned to the treatment group was still associated with a significant increase in the share of assignments completed prior to class (Table 5). It is likely that reviewing the material before lecture is an effective learning strategy for some but not all students, and individuals who scored relatively high on standardized college admission tests are already aware of and apply the methods that are most productive for them, so they do not benefit from being required to prepare before lecture. It is also plausible that the higher-ability students would have benefitted from materials that examined the material in more depth, whereas the content we provided was geared more toward students who had difficulty understanding the basic concepts.

The remaining two columns of Table 6 demonstrate that our results are essentially unchanged when we include instructor fixed effects or when we restrict the estimation to the final exam only. The GLS results corresponding to the IV models in columns 5 and 6 (not shown but available on request) show similar trends to those seen in column 2 of Table 6: not accounting for the endogeneity of preparation before class leads to downward bias for lower-ability students and upward bias for higher-ability students. This is again consistent with worse-performing lower-ability students preparing more for class but not benefitting enough from preparation to catch up to their higher-performing peers, and with positive selection into preparation for higher-achieving students. The estimated coefficients on other explanatory variables are similar to those in the ITT model and do not differ substantially across the various models presented in Table 6.

6. Conclusion

Previous studies such as Figlio, Rush, and Yin (2013) and Joyce et al. (2015) find that replacing face-to-face lectures with online lecture videos in an introductory economics course may have small negative impact on student learning. Our study differs, in that we explore how supplementing traditional lectures with online content enters the education production function and show that the timing is important. We find that requiring students to complete online reading assignments and watch videos whose content is similar to that presented in lectures is more effective when the work has to be completed prior to class.

While our setting—a Principles of Microeconomics course—is fairly specific, the implications of our study are quite general since many classes rely, either implicitly or explicitly, on students showing up to class prepared to learn the material. We show that most students in our Principles of Microeconomics course we study responded to required rather than recommended due dates, which is in line with research showing that individuals respond more strongly to external rather than self-imposed deadlines (Ariely and Wertenbroch 2002). We present two sets of results that can be useful in understanding how students learn economics and in developing best practices for teaching economics. The implication of the intent-to-treat results is that assigning prelecture materials improves exam performance for students in the middle two quartiles of the ability distribution, relative to making such materials available and requiring students to review them at any point during the semester. The local average treatment effect we estimate tells us that students in the middle of the ability distribution who review the material prior to lecture tend to do better on exams compared to students of similar ability who go over the same material but after lecture. In both cases, the improvement in performance corresponds to about a third of a letter grade.

Through the experiment we conduct, we investigate whether assignments whose objectives fall in the lower levels of the Bloom Taxonomy, remembering and understanding (Krathwohl 2002), can help students perform better in a microeconomics principles class. Since most students in our study completed the assignments at some point during the semester, our paper does not address the question of whether reviewing supplemental online materials is productive for students; rather, we show that the value of such materials is higher when instructors require students to use them prior to lecture. A limitation of our study is that we cannot distinguish clearly between the effects of requiring students to review the material prior to lecture from the effects of spacing out assignments throughout the semester versus assigning a large amount of work for the end of the term. It is important in future work to explore the exact mechanism through which reviewing the materials before class enhances student learning, whether it is through introducing the vocabulary that is to be used in lecture or familiarizing students with graphs. This would allow instructors to tailor the content of such assignments to their students' needs.

Appendix 1: List of Course Sections

Section	Type	Instructor ID	Taught by Ph.D. Student	Start time
1	Before	1	No	15:30
2	Before	1	No	15:30
3	End	1	No	9:30
4	End	1	No	9:30
5	Before	2	No	11:00
6	End	2	No	12:30
7	Before	3	Yes	14:00
8	Before	3	Yes	14:00
9	End	3	Yes	17:00
10	Before	4	Yes	14:00
11	End	4	Yes	11:00
12	End	4	Yes	15:30
13	Before	5	Yes	12:30
14	Before	5	Yes	15:30
15	End	5	Yes	14:00

References

Ariely, Dan, and Klaus Wertenbroch. 2002. Procrastination, deadlines and performance: Self-control by precommitment. Psychological Science 13(3):219–24.

Balaban, Rita A., Donna B. Gilleskie, and Uyen Tran. 2016. A quantitative evaluation of the flipped classroom in a large lecture principles of economics course. The Journal of Economic Education 47(4):269–87.

Becker, William E., and John R. Powers. 2001. Student performance, attrition, and class size given missing student data. Economics of Education Review 20:377–88.

Caviglia-Harris, Jill. 2016. Flipping the undergraduate economics classroom: Using online videos to enhance teaching and learning. Southern Economic Journal 83(1):321–31.

Emerson, T. L., and K. D. Mencken. 2011. Homework: To require or not? Online graded homework and student achievement. Perspectives on Economic Education Research 7(1):20–42.

Emerson, Tisha L. N., and Beck A. Taylor. 2004. Comparing student achievement across experimental and lecture oriented sections of a principles of microeconomics course. Southern Economic Journal 70(3):672–93.

Figlio, David, Mark Rush, and Lu Yin. 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):763–84.

Freeman, Scott, David Haak, and Mary Pat Wenderoth. 2011. Increased course structure improves performance in introductory biology. CBE-Life Sciences Education 10(2):175–86.

Freeman, Scott, Eileen OConnor, John W. Parks, Matthew Cunningham, David Hurley, David Haak, Clarissa Dirks, and Mary Pat Wenderoth. 2007. Prescribed active learning increases performance in introductory biology. CBE-Life Sciences Education 6:132–9.

Grodner, A, and N. J. Rupp. 2013. The role of homework in student learning outcomes: Evidence from a field experiment. The Journal of Economic Education 44(2):93–109.

Grove, Wayne A., and Tim Wasserman. 2006. Incentives and student learning: A natural experiment with economics problem sets. American Economic Review 96(2):447–52.

Haak, David C., Janneke HilleRisLambers, Emile Pitre, and Scott Freeman. 2011. Increased structure and active learning reduce the achievement gap in introductory biology. Science 332(6034):1213–6.

Harter, Cynthia, and John F. R. Harter. 2004. Teaching with technology: Does access to computer technology increase student achievement? Eastern Economic Journal 30(4):504–14.

Joyce, Ted, Sean Crockett, David A. Jaeger, Onur Altindag, and Stephen D. O'Connell. 2015. Does classroom time matter? Economics of Education Review 46:64–77.

Kirby, Kris N., Gordon C. Winston, and Mariana Santiesteban. 2005. Impatience and grades: Delay-discount rates correlate negatively with college GPA. Learning and Individual Differences 15(3):213–22.

Krathwohl, David R. 2002. A revision of blooms taxonomy: An overview. Theory into Practice 41(4):212–8.

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

Moravec, Martin, Adrienne Williams, Nancy Aguilar-Roca, and Diane K. O'Dowd. 2010. Learn before lecture: A strategy that improves learning outcomes in a large introductory biology class. CBE-Life Sciences Education 9:473–81.

Pennebaker JW, Gosling SD, Ferrell JD (2013) Daily Online Testing in Large Classes: Boosting College Performance while Reducing Achievement Gaps. PLoS ONE 8(11):e79774. https://doi.org/10.1371/journal.pone.0079774.

Seery, M. K, and R. Donnely. 2012. The implementation of pre-lecture resources to reduce in-class cognitive load: A case study for higher education chemistry. British Journal of Educational Technology 43(4):667–77.

Stanca, Luca. 2006. The effects of attendance on academic performance: Panel data evidence for introductory microeconomics. The Journal of Economic Education 37(3):251–66.

Swoboda, A., and L. Feiler. 2016. Measuring the effect of blended learning: Evidence from a selective liberal Arts College. American Economic Review 106(5):368–72.

Watts, Michael, and Georg Schaur. 2011. Teaching and assessment methods in undergraduate economics: A fourth national quinquennial survey. The Journal of Economic Education 42(3):294–309.

Watts, Michael, and William E. Becker. 2008. A little more than chalk and talk: Results from a third national survey of teaching methods in undergraduate economics courses. The Journal of Economic Education 39(3):273–86.