

The effect of supplementary video lectures on learning in undergraduate microeconomics

Melissa Famulari and Zachary A. Goodman*

University of California, San Diego

This version: October 2020

Abstract

The abstract goes here eventually.

*mfamulari@ucsd.edu and zgoodman@ucsd.edu. The authors thank the students who took intermediate microeconomics in the fall of 2018 and 2019 who consented to the use of their data for this study. We would also like to thank UC San Diego's Teaching and Learning Commons for anonymizing the data for analysis as well as the applied microeconomics group at UC San Diego for their help with the experimental design. This research was approved under UC San Diego's Human Research Protections Program (IRB approval 170886 in fall 2018 and 2019). The paper investigates the use of Intermediate Microeconomics Video Handbook (IMVH) video lectures by UC San Diego students, some of which were developed by one of the authors, in collaboration with UC San Diego and the UC Office of the President. UC San Diego currently owns the rights to distribute the IMVH. The videos lectures were provided to the subjects at no charge, nor did either of the authors have a direct financial interest during the study. As of fall 2020, one of the authors has a financial interest in future distribution of the IMVH outside of UC San Diego. The authors have no financial interest in the future distribution of the IMVH at UC San Diego.

1 Introduction

“You expect me to read the textbook? Ha!”

— Anonymous student

University students each spend tens of thousands of dollars annually on tuition and hundreds of hours in lecture and completing assignments, in large part, to learn. Instructors can improve how well students learn by employing pedagogical tools that have the greatest returns per unit time and financial cost. Despite the importance of comparing the effectiveness of different teaching technologies, little empirical work exists that estimates . In this paper we examine the impact of low marginal cost, video-based learning materials on exam scores in a large, intermediate microeconomic theory course.

The Intermediate Microeconomic Video Handbook (IMVH) at UC San Diego was designed to *supplement* lecture, not replace it, as an audiovisual version of a conventional course textbook. Part of the impetus for creating the IMVH was a discussion with a student who described her inability to read the course text, not because of poor reading skills, but because she did not find the text engaging enough to command her attention. We hypothesized that students, particularly Generation Z students who have had unprecedented exposure to electronic media, would find video materials more engaging and ultimately study more effectively or for more time than they would have if provided only conventional studying materials. Though students may use the IMVH more than the textbook, it had remained an empirical question whether the videos would ultimately improve learning outcomes.

We answer this question using a field experiment involving over 400 undergraduates enrolled in the same microeconomics course over two years. Only students who scored below the median on the first midterm were eligible for the experiment, since previous work and institutional knowledge suggests that students in the top half of the distribution would not benefit (and may even be harmed) from being forced to watch the videos. While the optimal experimental design for identifying average treatment effects would involve restricting access to the IMVH to only treated students, ethical considerations required that all students have access to the IMVH. Hence, we opted for an encouragement design in which treated students are induced to watch more videos than their control group peers through a grade-based

incentive, which more than doubled the number of videos watched by treated students. This experimental design permits identification of treatment effects local to those students induced by the encouragement to watch more videos.

We find that being assigned treatment (ITT) increased midterm and final exam scores by 0.18 and 0.17 standard deviations, respectively, and that the marginal hour of video watched increased exam scores (LATE) by 0.08 standard deviations. Although the confidence intervals are, admittedly, wide, the point estimates are statistically and economically significant: a student could increase their course letter grade by one step (e.g. from a B+ to A-) by watching XX hours of videos. Our estimates suggest that XX% of students in the control group who failed the course would have earned passing grades had they watched as many videos as their treated counterparts.

Although treated students performed better on course assessments, for assessing welfare it is important to determine whether time spent watching videos was additive to time spent engaged in other forms of studying. On one hand, if watching videos is more productive than the next best studying method, then the utility of requiring videos is unambiguously positive as students can substitute studying time towards the more productive option. On the other hand, if students must reduce time allocated towards leisure or studying for other classes so they can watch more videos, then the welfare implications are less clear and could be negative depending on the students' preferences.

We attempt to disentangle whether students spend more time studying or use their time more effectively by examining proxies for time use including class attendance, visits to a tutoring center (specific to this course), downloading materials from the course website, posting on the class discussion board, and reported time use from an in-class survey. Although our estimates are noisy, we find no statistically significant differences between treatment and control, and we can reject large decreases in take-up of other studying methods by treated students. Surprisingly, in nearly all cases, the point estimates suggest that the treatment group used studying methods beyond the videos at *greater* rates than did their control peers. We also investigate spillovers to other courses taken during the same academic term as the experiment and similarly find that treated students perform *better* than their control peers, which suggests that watching the videos likely did not dramatically reduce time spent

studying for other classes.

Neoclassical models of studying behavior assume that rational agents know their returns to studying using the methods available to them and allocate the optimal amount of study time to each method given their utility function, which is increasing in leisure and grades and decreasing in time spent studying.¹ Since college instructors typically have little knowledge about their students' other classes or aspects of their lives that may have large payoffs in the labor markets (leadership positions in student organizations, internships, etc.) or marriage market (when else does one live in a veritable city of people of similar age, ability and interests?)², the neoclassical models suggest that instructors cannot make their students better off by intervening in their studying allocation decisions.

However, many would agree that the "raison d'être" of higher education is to teach students how to learn. There is significant evidence from psychology that many college students do not know how to learn effectively³. Universities often fund "Teaching and Learning Centers" or "Academic Skills Centers", part of whose mission is to help undergraduates study more efficiently⁴. We posit that for many students, a key assumption of the neoclassical model does not hold: that students possess correct information about the returns to studying of different studying methods. Instead, we offer the alternative hypothesis that students supply a quantity of study time that is optimal given their information constraints. They choose study methods and quantities that are suboptimal relative to those they would have picked in a full information setting. Hence, an intervention by an entity that has more

¹Oettinger (2002), "The effect of nonlinear incentives on performance: Evidence from Econ 101," *The Review of Economics and Statistics*, 84(3): 509–517, provides evidence that student effort responds rationally to grade incentives. Across 1200 students in an economics class with absolute grading standards, he finds evidence of bunching just above the letter grade cutoffs and student performance on the final exam is higher if the student is just below a grade threshold

²Attanasio and Kaufmann (2017) "Education choices and returns on the labor and marriage markets: Evidence from data on subjective expectations," *Journal of Economic Behavior & Organization* 140 (2017) 35–55

³See, for example, Pashler, Rohrer, Cepeda, Carpenter (2007). "Enhancing learning and retarding forgetting: Choices and consequences," *Psychonomic Bulletin and Review* 2007, 14, 2, 187-193; Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M.J., and Willingham, D. T. (2013). "Improving students' learning with effective learning techniques promising directions from cognitive and educational psychology." *Psychological Science in the Public Interest*. Afton Kirk-Johnson, Brian M. Galla, Scott H. Fraundorf (2019). "Perceiving effort as poor learning: The misinterpreted-effort hypothesis of how experienced effort and perceived learning relate to study strategy choice" *Cognitive Psychology*, Volume 115

⁴All nine University of California campuses have one. Some others in the US include Dartmouth's Academic Skills Center, Michigan's Center for Research on Teaching and Learning, UNC's Learning Center, and Yale's Teaching and Learning Center

information about returns to studying (i.e. an instructor) can be utility enhancing.

A competing model consistent with the observed difference in behavior between treatment and control is of the behavioral sort in which students plan to study more than they end up studying when the time comes. Indeed, survey and experimental data suggest that many students study less than they report they “should” and finish the term with grades lower than what they had anticipated they would earn at the start of the term⁵. This phenomenon is consistent with two-self models in which a person’s “planner” self, the one who desires high grades at the expense of leisure, is at odds with her “doer” self, the one who must choose between immediately gratifying leisure and delayed gratification from higher grades⁶.

One important (and testable) difference between the incomplete information model and the two-selves model is what happens after exogenous incentives to watch videos are removed. While the former predicts that students exposed to treatment will continue watching videos given their new knowledge of the highly productive studying technology, the latter predicts that the students will return to their lower baseline levels of video watching as their doer selves no longer have a commitment device reducing the temptation of immediately gratifying leisure. We examine video watching behavior during the term following the experiment in the subsequent microeconomics course and find that treated students watch significantly more videos than their control classmates, consistent with the incomplete information model. Collectively, we interpret these findings as strong evidence that requiring studying tools known by the instructor to be effective is utility enhancing for students who make manifest their limited knowledge of how best to study, perhaps through poor performance on an early stage assessment.

The rest of the paper is organized as follows. Section 2 provides background on existing related literature. Section 3 describes the experimental design. Section 4 presents the results of the experiment, and Section 5 discusses those results. Section 6 concludes.

⁵Ferrari, J.R. “Psychometric validation of two Procrastination inventories for adults: Arousal and avoidance measures.” *J Psychopathol Behav Assess* 14, 97–110 (1992). Patricia Chen, Omar Chavez, Desmond C. Ong, Brenda Gunderson (2017) “Strategic Resource Use for Learning: A Self-Administered Intervention That Guides Self-Reflection on Effective Resource Use Enhances Academic Performance” *Psychological Science*, Vol. 28, Issue 6, 774-785

⁶see review paper, Adam M. Lavecchia, Heidi Liu, and Philip Oreopoulos (2016), “Behavioral Economics of Education: Progress and Possibilities” *Handbook of the Economics of Education*, Volume 5, Pages 1-74)

2 Background

Worth it to make a distinction between start of the course design (textbooks, # exams, etc.) vs after student has revealed they are a low performer? Carroll and Kurlander (2020) make distinction between student interventions vs instructor interventions.

Previous research has examined the effects of inducing students to alter their inputs into the educational production function. Dobkin, Gil and Marion (2010) required students scoring below the median on the midterm to attend class. The attendance policy led to a 28 percentage point reduction in skipping class at the threshold. Using a regression discontinuity approach, they find that a 10 percentage point increase in attendance results in a 0.17 standard deviation increase in the final exam score with no effect on the grades in other classes taken the same quarter. Effects of quizzes.

3 Experimental Design

We conducted the field experiment during an intermediate microeconomics course over two quarters, fall 2018 and 2019, taught at a large, highly selective public research university. At this institution, intermediate microeconomics is a three-term sequence required for all students majoring in Economics, joint Economics-Mathematics, and Management Science. The experiment was conducted in the first course of the sequence, though we can observe grades in the second course of the sequence.

Students were informed about the experiment during the first lecture and in the syllabus, which was provided to all students via the course webpage. At any time during the quarter, students were permitted to opt out of having their anonymized data included in the analysis, which would not begin until grades had been posted and data had been anonymized⁷. Students below the age of 18 at the start of the course as well as pupils enrolled in the course via an online extension program were removed from the analysis dataset.

The experiment began four weeks into the term following the first midterm exam, and as

⁷Students could opt out via a form handled by the Teaching + Learning Commons (T+LC) so that neither the instructor nor research team could observe which students decided to opt out. The T+LC also anonymized the data for the research team before any analysis was conducted.

Table 1: Grade scheme by treatment arm. *Control* represents same grade scheme as *Above median*. Differences between the two grade schemes in red.

Assessment	Incentive	Control
>40 videos	4%	0%
Midterm 1	18%	22%
Midterm 2	22%	22%
Final Exam	50%	50%
Math Quiz	1%	1%
Best 5 of 6 Quizzes	5%	5%
Total	100%	100%

such, only students who took the first midterm were included in the experiment. Following the first midterm, students were assigned to one of three treatment arms: “Incentive”, “Control”, or “Above median”. Students who scored below the median on the first midterm were randomly assigned to *Incentive* or *Control* whereas *Above median* includes all other students. Random assignment was performed using pairwise stratification conditional on midterm score and year cohort (specific detail on treatment assignment can be found in the Appendix). Students in the *Incentive* arm received a grading scheme that encouraged watching videos during the rest of the term whereas *Control* and *Above the median* received the standard grading scheme that does not directly reward watching videos. The two different grading schemes are outlined in Table 1. Specifically, the *Incentive* grading scheme requires that at least 40 of 48 eligible videos in the IMVH be watched to earn 4 percentage points towards the students’ final grades⁸. Notably, the 4 percentage points comes at the expense of the first midterm score, which had already occurred at the time of treatment assignment. Hence, we isolate the video incentive as the sole difference between treatment arms, giving us more confidence that the exclusivity assumption of our encouragement design holds.

which and there are two unique features of this class. First, many non-majors take the class typically to either satisfy general education requirements or to explore majoring in Economics. Thus there are many students at the margin of majoring in economics in the class and so an important outcome is the likelihood the student takes the second class in the sequence. The other unique feature of this class is the large fraction of transfer students,

⁸Watched in standard speed, 40 videos would require students to spend between 5.5 and 7.1 hours, depending on the length of videos chosen (on average 9.7 minutes in length each). Watching all incentivized videos in standard speed would require just shy of eight hours.

for whom the class is not only their first experience with upper division coursework, but typically the first time taking a class under the quarter system (community colleges in the state are on the semester system), and their first class at a 4-year research university. Thus, we expect transfer students to have greater information problems.

While the experiment was conducted in the first class of the sequence (100A), we also have data from the second class in the sequence (100B). Fortunately, one instructor taught all four of the 100A classes (one of the authors) and another instructor taught all four of the 100B classes. Both instructors created half of the videos lectures for their course.

Enrollments are high enough that the 100A and 100B instructors both taught two classes back-to-back each quarter but offered all exams at a common time out-of-class. So we treat the two classes in a quarter the same but account for different years in the empirical analysis.

The 100A course had an identical structure across all quarters and years. In addition to the textbook and the option to attend a live lecture at either 11-12:30 or 12:30-2:00, students had access to weekly one-hour discussion sections run by graduate TAs who were all Economics PhD candidates (including one of the authors), a tutoring lab staffed by both the graduate TAs (in lieu of office hours) and top undergraduates (some earning course credit for learning how to teach economics and others hired by the Department to help cover the tutoring lab hours, M-Th 5:30-8:30 and Sunday 4-8pm), weekly supplemental instruction sessions offered by an undergraduate majoring in Economics and trained by the university in supplemental instruction, a discussion board (monitored by the instructor as well as the grad and undergrad TAs), four years of the instructor's old exams (no answers provided), weekly ungraded problem sets (with detailed answers), graded online quizzes on Mondays of most weeks that did not include an exam, and a video handbook containing 220 short-ish videos covering the entire 100ABC sequence. This video handbook was created by the two instructors teaching 100A and 100B as well as four other faculty members who also regularly teach in the intermediate microeconomics sequence. Assessment in the class was as follows:

Prerequisite math quiz (in-class, second meeting) 10 points

Top 5 of 6 online quizzes 50 points

Midterm 1 (week 3) 220 points

Midterm 2 (week 7) 220 points

Final Exam (Saturday after week 10) 500 points

Students scoring below the median on the median on midterm 1 were randomly assigned to a "Videos Required" group where their midterm 1 score was down-weighted to 180 points and 40 points (all or nothing, so no partial credit) was put on watching 40 of the 47 remaining videos. Students were told that final letter grades would not be affected by being in the experiment, so the two key experimental outcomes are the scores on the second midterm and on the final exam. Following [cite], we randomized the students into the experiment by sorting the students by the first midterm score, creating ordered student pairs and then choosing one of students in the pair randomly.

Video "watching" was based on opening the video, leaving it open for at least half the video length, and recording e-mail address on link at end of the video. Right after random assignment Twice during the quarter we let students in the experiment

In particular, we use an assessment (exam) to identify students most likely to be struggling with meeting the learning objectives of the course. We require a random sample of students who score below the median on the first midterm to partake in a required (for one's grade) study strategy consisting of watching at least 40 course-relevant supplementary videos. The videos were freely available to all students in the class and the instructor committed to having the course grade distribution identical across treatment and controls. To keep the weights on the second midterm and final exam, our primary outcome measures, the same across the treated and control groups, the first midterm was down-weighted for the treated group and the weight was put on watching 40 videos. Nearly all students in the treated group watched the 40 videos and treatment led to a X percentage point increase in video views relative to the control group. We find the grading policy led to an

We are primarily interested in the causal effect of watching videos on exam grades. The average causal effect of watching videos can be modeled using the potential outcome framework or Rubin Causal Model (Imbens and Rubin, 2015):

$$\tau = E[Y_{it}(1) - Y_{it}(0)] \tag{1}$$

where τ is the average causal effect of watching videos and $Y_{it}(1)$ and $Y_{it}(0)$ are the potential outcomes (e.g exam scores) for a student i in year t who does and does not watch videos, respectively. We can observe treatment assignment for each student, $Z_{it} \in \{0, 1\}$, as well as the observed outcome $Y_{it}(Z = z_{it})$ and a vector of pretreatment covariates X_{it} .

A student’s decision to watch videos is endogenous, so we must rely on an exogenous instrument to calculate an unbiased estimate of τ . By randomly assigning treatment that induces significantly greater video watching, the treatment indicator Z_{it} satisfies the instrument validity and relevancy conditions required to estimate τ using an instrumental variable approach (citation). We calculate an estimate of τ using a two-stage least squares approach:

$$v_{it} = \alpha Z_{it} + f(X_{it}) + e_{it} \quad (2)$$

$$y_{it} = \tau \hat{v}_{it} + g(X_{it}) + u_{it} \quad (3)$$

where \hat{v}_{it} is instrumented videos watched estimated by Equation 2, $f()$ and $g()$ are generic functions through which X_{it} affects v_{it} and y_{it} , respectively, and e_{it} and u_{it} are model residuals assumed to be mean-zero conditional on observables. $\hat{\tau}$ is the estimate of the local average treatment effect (LATE) of watching videos local to those students induced by treatment to watch videos.

Under the assumptions of independence, monotonicity, and non-interference, $\hat{\tau}$ is an unbiased estimate of the LATE (Angrist and Imbens, 1995). Independence assumes that outcomes (e.g. grades) are only impacted by treatment through watching videos. This assumption could be violated if, for example, telling a student she is treated were to give her more confidence on subsequent exams during the quarter. Monotonicity, sometimes referred to as the “no defiers assumption”, is required because of two-sided noncompliance and requires that students assigned treatment are weakly more likely to watch videos than if they were assigned control. A violation of this assumption could occur if students get utility from rebelling against their assigned grade scheme. Non-interference, also known as the Stable Unit Treatment Value Assumption (SUTVA), assumes that each student’s outcome depends only on their own treatment status and not the treatment status of their peers.

Violations of SUTVA may include control students benefiting from having treated students in the same class (and perhaps studying together).

Although we believe excludability⁹ and monotonicity¹⁰ are reasonable assumptions, we have more pause about the non-interference assumption because of the potential for spillovers between students in the same class. If we had unlimited resources, a more robust experimental design would assign treatment at the class (or coarser) level, reducing the chance for interactions between treated and control students. However, given our resource constraints, assigning treatment at coarser levels would have resulted in insufficient statistical power to detect large effect sizes. Hence, we proceed acknowledging the potential for spillovers between students. We hypothesize that spillovers likely bias our estimates of the treatment effect *downwards* as we believe control students are more likely to benefit from having well-studied peers than they are to lose from, for example, having peers too busy watching videos to join a study group.

We also estimate the average causal effects of being assigned to the *Incentive* arm, or average Intention To Treat (ITT) effects. These are ITT estimates and not average treatment effect estimates because of non-compliance: some students in the treatment arm do not watch videos and some students in the control arm do watch videos. However, since the incentive itself in our setting is representative of how future instructors may require their students to use the IMVH, the ITT estimates are relevant for an instructor interested in deciding whether to require her students watch videos or simply make them available with no grade incentive.

Following the methods proposed by Robinson (1988) and more recently Wager and Athey (2018), our primary estimating equation is the partially linear model:

$$Y_{it} = \beta Z_{it} + f(X_{it}) + \epsilon_{it} \tag{4}$$

where Y_{it} is an outcome of interest (e.g. videos watched or test scores) for student i in

⁹While this assumption is not testable, we took care in the experimental design to make the treatment and control arms as similar as possible except for the grading schemes

¹⁰Though not testable directly, one testable implication of monotonicity is that the cumulative distribution function of videos watched for each treatment arm should not cross. Indeed, in Appendix Figure ?? shows that the two CDFs do not cross.

year t , $Z_{it} \in \{0, 1\}$ is a treatment indicator, $f()$ is a generic function through which X_{it} , a vector of controls, affects Y_{it} , and ϵ_{it} is an unobserved residual. Following the Neyman-Rubin potential outcomes framework, $Y(Z = 1|X) - Y(Z = 0|X)$

To identify the causal effect parameter of interest, β , we make three assumptions: 1) *treatment status unconfounded given controls*: $Z_{it} \perp\!\!\!\perp (Y_{it}(Z = 1), Y_{it}(Z = 0)) \Big| X_{it}$; 2) *Excludability*: 3) *Non-interference/Stable Unit Treatment Value Assumption (SUTVA)*:

4 Results

5 Discussion

6 Conclusion

Students come to college, in part, to learn how to learn. In large classes it is not possible to meet individually with every student to provide individual advice on effective studying and students who likely would benefit the most from such advice may be the least likely to come ask the instructor or TA for help with how to pass the class. This suggests "learning interventions" may be particularly effective in classes taken in the first quarter/semester, the first upper division classes in the major, and/or large classes. Further, transfer students, who simultaneously take their first class at the university, first large class, and first upper division class may particularly benefit from such interventions. Finally, students at the margin of success in a major may benefit

In this paper we

References

Angrist, J. D. and G. W. Imbens (1995). "Two-stage least squares estimation of average causal effects in models with variable treatment intensity". In: *Journal of the American statistical Association* 90.430, pp. 431–442.

- Athey, S. and G. W. Imbens (2017). “The econometrics of randomized experiments”. In: *Handbook of economic field experiments*. Vol. 1. Elsevier, pp. 73–140.
- Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
- Robinson, P. M. (1988). “Root-N-consistent semiparametric regression”. In: *Econometrica: Journal of the Econometric Society*, pp. 931–954.
- Wager, S. and S. Athey (2018). “Estimation and inference of heterogeneous treatment effects using random forests”. In: *Journal of the American Statistical Association* 113.523, pp. 1228–1242.

7 Appendix