The effect of supplementary video lectures on learning in intermediate microeconomics

Melissa Famulari and Zachary A. Goodman*

University of California, San Diego

This version: October 2020

Abstract

The abstract goes here eventually.

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 current 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 is an empirical question whether the videos 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 induced 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 percent 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 determining welfare it is important to identify where the time watching videos came from: leisure time, working, student organizations, studying for other classes, studying for current class using other methods, etc. 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 treated students spent more time studying or used 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 study methods by treated students. Surprisingly, in nearly all cases, the point estimates suggest that the treatment group used study methods beyond the videos at *greater* rates than did their control peers. Though estimates are noisey, we find no significant differences in reported lesire time across treatment and control. Finally, we 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.

Finally, we attempt to distinguish between two models of student learning which could explain why the video watching inducement improved student exam performance: an incomplete information model where students do not know how to study effectively versus a two-selves model where students like good grades but dislike studying are 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 a relatively 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 our findings as strong evidence that requiring studying tools known by the instructor to be effective is utility enhancing for students who manifest their limited knowledge of how best to study, perhaps through poor performance on an early stage assessment. Finally, we provide suggestive evidence that students found the IMVH to be a relatively effective study method by examining video watching across the treatment and controls in the next class in the sequence. We hypothesize that treated students will continue to watch videos at a higher rate than the controls, even in the absence of a grade incentive, if the treated students found the videos to be a relatively more effective study tool. Our estimates suggest that treated students watched X percent more videos in the next class in the sequence.

The rest of the paper is organized as follows. Section 3 provides background on existing related literature. Section 4 describes the experimental design. Section 5 presents the results of the experiment, and Section 6 discusses those results. Section 7 concludes.

2   Models of Studying Behavior

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. Oettinger (2002) provides evidence that student effort responds rationally to grade incentives.[1] Since college instructors typically have no knowledge of the student's utility function and so do not know student preferences over other classes and aspects of their lives that may have large payoffs in the labor market (leadership positions in student organizations, internships, networking, etc.) or marriage market (when else does one live in a veritable city of similarly aged people with similar ability and interests?)[2], neoclassical models suggest that instructors cannot make their students better off by intervening in their study decisions.

However, in addition to teaching specific skills, many would agree that the "raison d'etre" of higher education is to teach students how to learn. There is evidence from psychology that college students do not know how to learn effectively.[3] Universities often fund "Teaching and Learning Centers" or "Academic Skills Centers," part of whose mission is to help undergraduates study more efficiently.[4] We posit that for many students, a key assumption of the neoclassical model does not hold: that students possess accurate information about the returns across studying methods. Instead, we offer the alternative hypothesis that students supply a quantity of study time that is optimal given their information constraints.

---

[1]Oettinger (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.

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

[3]See, 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

[4]All 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

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

A third model is a behavioral one 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 term[5]. Recent empirical evidence suggests that setting tasked-based goals help improve college student performance.[6] 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 grades.[7]

Suppose an instructor uses a grade incentive to encourage the use of a time consuming educational input, say, watching instuctional videos (or attending class, reading the textbook, taking quizzes, etc.) Assume the partial derivative of final exam score with respect to watching videos is positive.[8] Also assume that the grade incentive is effective and students do watch more videos. What are the testable implications of the three models?

In the neoclassical model, student utility will be lower with the instructor's incentive to watch videos. A sharp prediction is that, in the absense of the incentive, student video watching will revert to pre-incentive levels. It is possible to find lower exam scores if students keep total study time for the class constant and substitute towards videos and away from

[5]Ferrari, 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

[6]Clark, Damon, David Gill, Victoria Prowse, and Mark Rush (2020) "Using Goals to Motivate College Students: Theory and Evidence From Field Experiments" The Review of Economics and Statistics 2020 102:4, 648-663

[7]see 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)

[8]We expect non-native speakers to particularly benefit from the videos because of the technology itself (e.g. ability to replay videos, change speed and having closed captions). We test for this interaction in the results below.

more productive inputs (study groups, attending class, reading the textbook, etc). It is also possible that exam scores will increase if the incentive to watch videos increases total study time for the class. If total study time in the course increases, where did the additional time come from? Study time in other classes? Then the incentive may increase performance in the class at the expense of grades in other classes. Time devoted to student organizations? Then the incentive may reduce the likelihood of being in a leadership position. In the empirical section, we test for the effects of being induced to watch videos on exam performance as well as on other uses of the student's time incuding lecture attendance, visits to a class-specific tutoring lab, use of a class discussion board, downloads from the class web page, and student survey responses regarding study and leisure time.

In the poor student information model where the instructor has perfect information, the incentive to watch videos will increase exam performance if student study time in the course does not fall. While unlikely, students with poor information may increase video watching but reduce other, less effective educational inputs so much that exam performance falls. A sharp prediction is that video watching will continue at the incentivized level even in the absence of the incentive as students will have learned about an effective study tool. In the empirical section, we test whether treated students watched more videos in the second course in the sequence. We also examine whether the LATE is greater for transfer students who are likely to have more information problems because they are taking their first class at a 4 year, R1 university and their first class under the quarter-system.

Finally, in the behavioral model, the instructor's inducement helps students stick to study plans. Again, as long as total study time does not fall, the inducment will increase exam performance. However, there is nothing about the videos per say that leads to the increased exam performance. In the absence of the inducement, we expect video watching to fall to pre-inducment levels.

3   Related Literature and Contributions

The intervention we study, inducing students to watch pre-recorded instructional videos, almost exclusively involves student time [9]. Since the IMVH is the first videobook to our knowledge, ours is the first to examine the effects of access to a video book. However, students have many alternatives to help them learn the course material, including attending class, reading the textbook, doing homework, and taking quizzes, and there is a large empirical literature examining the effects of each. For each learning method, a positive correlation between course performance and use of the learning method is not enough to establish causality because of unobserved student characteristics, such as motivation, that are positively correlated with the use of a learning method and class performance. As a result, we focus on empirical studies that use experiments and quasi-experiments.

Dobkin, Gil and Marion (2010) analyze a policy where lecture attendance was voluntary before the midterm, but after the midterm, students scoring below the median were required to attend class. The policy affected 352 students taking three classes, two intermediate micro and one econometrics class. The policy led to a 36 percentage point increase in post-midterm attendance at the threshold. Using a regression discontinuity design, they find that a 10 percentage point increase in overall attendance results in a 0.17 standard deviation increase in the final exam score. They find no effect of the attendance policy on grades in other classes taken the same quarter.

Effects of quizzes.

Other researchers have investigated using technology to improve learning. **nbbm2020** conduct an experiment in Botswana during the COVID-19 pandemic and find that text messages and phone calls deployed as low cost, scalable learning technologies improved test scores by 0.16 to 0.29 standard deviations.

This study adds to this body of research by studying the effectiveness of an educational innovation: a video textbook. The Intermediate Microeconomics Video Handbook (IMVH) is a collection of 220 shortish videos that cover the material for a year-long intermediate

---

[9]As with textbooks, instructor time is used if students ask questions about differences between lecture and the videos or textbook. It can be particularly time consuming if the answer to an exam question differs across lecture, videos or textbook

microeconomics class. Most topics were covered by two videos: one video introduces the concept more intuitively with verbal explanations and graphs and the other video has the more formal, calculus-based presentation. The videos were created in 2014 by six faculty members with professional videographer and production support. Many videos were created using an innovative presentation technology, the "learning glass," where the instructor uses neon markers to write on a large sheet of glass, with lights embedded along the glass edge to make the colors pop. The remaining videos are a PowerPoint presentation that faculty could write on and the instructor is superimposed next to the PowerPoint explaining the concept. Each video was closed captioned and the captions are searchable. Each video also includes time stamps so students can quickly find the material they want to learn or review. These videos were organized into a book with both a table of contents and an index.

The IMVH is not an e-book which typically takes a physical book and digitizes it. The IMVH differs from a traditional textbook because intructors explain, graph and derive mathematical results in the same way one would in a lecture. The IMVH differs from lectures because students cannot stop the presentation, ask questions and get answers in real time. The IMVH also differs from lectures because students can rewatch when they are confused, speed through material they understand or skip to parts of the video. The IMVH is closed-captioned which may be particularly useful when the instructor and/or the student is a non-native speker

4   Experimental Design

We conducted the field experiment in four intermediate microeconomics courses, two in fall 22018 and two in fall 2019. The university is a large, diverse and selective public research four-year university. At this institution, intermediate microeconomics is a three-term sequence required for students majoring in Economics. The experiment was conducted in the first course of the sequence. We also have grades and video viewing 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. A hard copy of the consent form

was given to the students on the second day of class. 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 anonymized[10]. 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 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 grades[11]. 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,

---

[10]Students 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.

[11]Watched 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.

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%

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.

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 to be in the experiment.

Video "watching" was based on opening the video, leaving it open for at least half the video length, and clicking a link at the end of the video which took students to a Qualtrics survey where they had to record their e-mail address on link at end of the video. Right after random assignment we surveyed students to make sure they knew which group they were in. Twice during the quarter we let students in the experiment know how many videos they had watched

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 (**ir2015**):

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

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:

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

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

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 (**ai1995**). 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 refered 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 con-

trol. 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 excludability[12] and monotonicity[13] 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 **robinson1988** and more recently **wa2018**, our pri-

---

[12]While 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

[13]Though 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.

mary 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 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*: $Z_{it} \perp\!\!\!\perp \left(Y_{it}(Z = 1), Y_{it}(Z = 0) \right) \Big| X_{it}$; 2) *Excludability*: 3) *Non-interference/Stable Unit Treatment Value Assumption (SUTVA)*:

5 Results

6 Discussion

7 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 technoques 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.

Which type of classes? This suggests "learning interventions" may be particularly effective for typically classes taken in the student's first quarter/semester, the first upper division classes in the major, and/or large classes.

What type of students? Freshman. Also 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 capped major may benefit from an intervention to help them meet the grade requirement. UCOP study examined students who just barely missed the grade cutoff into an economics major and find

their early career earnings are substantially lower than students who just met the cutoff.

Our specific learning intervention: intermediate micro, a notoriously challenging course across all economics programs. May also particulalry benefit non-native speakers due to closed captioninng and ability to slow and/or replay the video.

We examine the effectiveness of an educational innovation, a video handbook, where 220 (mostly) short instructional videos were organized into a book. Even if the marginal cost of another video view is nearly zero, instructors may have concerns about making a new resource available to students due to connerns about substitution away from more productive inputs. We focus on students who demonstrate they may benefit from an educational intervention by scoring below the median on an early assessment. Of course, these students could also be the most hurt by the intervention if the reason for doing poorly on the early assessment was high value of time in alternate uses.

Our field experiment is unable to determine whether the videos uniquely helped the treated students or whether any intervention that induced additional time spent on the course would have been effective. We do have suggestive evidence that the videos were uniquely effective: treated students in the next couse in the sequence, where there was no grade inducement to watch videos, continued to watch the videos at a higher rate than the control students.

8 Appendix