MOOCs as a Massive Research Laboratory

Ignacio Martinez
Tucumán, Argentina

M.A. Economics, University of Virginia, 2010
Lic. Economics, Universidad Nacional de Tucumán, 2008

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of Doctor of Philosphy

Department of Economics

University of Virginia
July, 2014

Abstract

Massive Open Online Courses (MOOCs) offer students a new way to learn, and researchers a new laboratory for studying learning. Education researchers have struggled to understand how students learn and what helps them achieve more. The "Big Data" environment enabled by the technology used to deliver MOOCs provides an unprecedented opportunity to get inside the "black box" of student learning. On the other hand, the sheer scale of MOOCs combined with the extraordinary dimensionality of the process and outcome data also imply substantial challenges for monitoring outcomes.

In the first chapter, co-authored with Paul Diver, I explore the opportunities and challenges that MOOCs are generating for research. A wide variety of topics related to pedagogical methods and student incentives lend themselves to research using MOOCs; throughout the chapter, I discuss lessons that can be gained both from observational comparisons and especially from the opportunity to run experiments on randomly chosen groups of students. I start by discussing dropout rates and study how students who decide to drop out are different from those who continue in the course. I then discuss class forums and video lectures and how interaction with this material is correlated with achievement. After that, I explore the strong correlation between procrastination and achievement and the implications for course design. I also examine the role of certification offered by MOOCs and how certification options can affect choices and outcomes. Finally, I examine the potential of linking data across courses and the opportunities and challenges of working with data that originates in surveys of MOOC participants. All of these research opportunities offer Big Data challenges as well which have to be addressed with parallel computing.

In the next three chapters, I present results from randomized experiments that I run in the big-data environment offered by MOOCs. Both experiments take the form of "nudges", emails that simply provide certain information that may suggest changes in behavior to students. In the second chapter, I evaluate the impact of providing students with information about their performance relative to their classmates. I run a randomized experiment in the context of a Coursera MOOC, assigning students to either one of two potential treatments. The first, framed positively, describes what fraction of his classmates a student outperformed. The second, framed negatively, describes what fraction of his classmates a student underperformed. I find evidence that students respond to this informational nudge and that framing matters. Students who were doing relatively poorly respond to the negative treatment with more effort, and this effort translates, in some cases, into higher achievement. On the other hand, students who were doing relatively well respond to the positive treatment. As an example, the average student in the control group, among those who did not have a perfect score on the first quiz before the intervention, was ranked in the 31.6 percentile of the class on the third quiz, while the average student after receiving the negatively framed treatment was in the 40.5 percentile.

The third chapter identifies the causal effect of procrastination on achievement in a MOOC. I use two approaches, instrumental variables (IV) and a randomized control trial. I show that rain and snow affect when a student takes a quiz, and, therefore, can be use as an IV. I find that taking the course first quiz on the day it is published, rather than procrastinating, increases the probability of course completion by 15.4 percentage points. For the randomized control trial, I send an email (directive nudge) encouraging a randomly selected group of students to procrastinate less. I find that the effects are heterogeneous across countries, suggesting that it may be advisable to customize nudges to country characteristics. As an example of the magnitude of the effects, Germans assigned to the treatment group were 167% more likely to obtain the course certificate while there is no effect for Americans. This shows that very low-cost intervention can increase student achievement. This online experiment may provide valuable lessons for traditional classrooms.

The last chapter, co-authored with Louis Bloomfield and Sarah Turner, shows that a MOOC can serve as a complement to a brick-and-mortar introductory physics course. I randomly assigned two thirds of the students to receive a small monetary incentives to enroll in a MOOC. Half of the treatment group received a $10 Amazon gift card simply for enrolling in the MOOC and responding to our email. The other half received a $50 Amazon gift card if they responded to the email and obtained an 80% final score in the MOOC. Using these monetary incentives as instruments for enrolling in the MOOC, I show that MOOC enrollment significantly improves performance in the brick-and-mortar classroom.

Taken as a whole, these essays describe how we can use MOOCs to learn more about the student achievement production function. Additionally I show that very low-cost interventions can nudge students into changing their behavior and improve their achievement. As technology advances, MOOCs providers collect more data on how their users interact with their platforms. For example, better data on time use. Additionally, technology will soon allow for more complex interventions than the simple ones in these essays. This will ultimately allow us to determine optimal course design, and learn more about personalized learning.

Acknowledgements

I could not have written this dissertation without the support and guidance of my advisors at the University of Virginia. I will forever be indebted to Sarah Turner and Leora Friedberg. I also received valuable comments from Benjamin Castleman, Beatrice Boulu-Reshef, John Pepper, Denis Nekipelov, and Steven Stern.

I benefited, both professionally and personally, from relationships with many of the friends I met in the UVa graduate program, including Charlie Murry, Catherine Alford, Heidi Schramm, Jeff Hulbert, Felipe Benguria, Andrew Barr, Matt Goldman, and especially Patten Mahler.

I would also like to thank Coursera for providing me with the data, especially Tom Do, James McMath, Eli Bildner, and Andreina Parisi-Amon. On the UVa side of the data collaboration I am grateful to Kristin Palmer, and Professors Lou Bloomfield, Mike Lennox, and Philip Zelikow. Finally, for their invaluable computational support, thanks to Katherine Holcomb and Jackie Huband from the Tiger Team.

Much of the research in this dissertation was done while receiving financial support from the Bankard Fund for Political Economy, the Jefferson Trust, and the office of the vice-president for research.

Finally, I would like to thank Carina, Gaby, Leo, and Tomy. This accomplishment would not have been possible without their encouragement and love.

This dissertation is dedicated to

Leo

Contents

Chapter 1

Opportunities and Challenges.

Coauthored with Paul Diver.

1.1 Introduction

Public interest in Massive Open Online Courses (MOOCs) has been growing over the past few years. Searching for MOOCs in The New York Times returns almost 400 articles, starting with only 2 dated back to 2011.[1] Figure 1.1 shows search interest over time using data from Google. In 2013, William Bowen, former president of Princeton University and the Mellon Foundation, argued that online learning is here to stay, see Bowen (2013). He attributed this to a combination of three factors. First, technological advances have reduced storage cost, and improved Internet speed and availability. Second, students are embracing all things digital and expect to communicate in this way. Third, the cost of traditional higher education keeps increasing over time.

We argue that MOOCs not only offer students a new way to learn but also offer researchers a new laboratory for understanding learning. Education researchers have struggled to know how students learn and what helps students learn better. The "Big

[1] As of March 19^{th} 2014.

Data" environment enabled by the technology used to deliver MOOCs provides an unprecedented opportunity to get inside the "black box" of student learning. On the other hand, the sheer scale of MOOCs combined with the extraordinary dimensionality of the process and outcome data also imply substantial challenges for monitoring outcomes.

In this paper we analyze learning outcomes in two MOOCs offered as a partnership between Coursera and the University of Virginia. *Foundations of Business Strategy* (FBS) is a 6-week course that had 84,377 students initially enrolled. *The Modern World: Global History since 1760* (MWH) is a 14-week course that had 46,575 students initially enrolled. Coursera records every single click a student makes, with the meta-data that tell us what they are clicking and when. We are able to observe every individual answer for every attempt they made on each quiz. We know how many forums threads they created and how many they read. We are able to see when they clicked play, pause, or fast-forward during a video lecture. These rich data generate great opportunities for research, but also great challenges. For example, the file that contains the click stream for MWH is a 16GB JSON file. Each line of that file contains single click meta-data that has to be processed in order to transform it into useful information. The rest of the data, including on quizzes, is provided in an SQL dump with many tables that also have to be processed. In this paper, we discuss the challenges of manipulating these data and how parallel computing offers some resolutions.

In order to give a sense of the types of learning outcomes that we can analyze using MOOCs, and that are much more difficult to get a handle on in bricks-and-mortar classrooms, we discuss lessons that can be gained both from observational comparisons as well as from the opportunity to run experiments on randomly chosen groups of students. We start our analysis by comparing choices and outcomes for students who drop out after the first quiz and students who continue to the second. We show that students who drop out interact less with forums and videos than students who continue. Using data from FBS, in which students can take the quizzes multiple times, we show that students who read the forums between their first and best attempt improved their grade more than those who did not read the forums. Similarly, students who went back to the video lectures improved their grade more than those who did not. Because all interactions with the course are time-stamped, we are able to see how much time passes between when a quiz is posted and when a student takes it the posting of a quiz and when a student solved it. We find a strong negative correlation between procrastination and grades. For example, MHW students who obtained the certificate with a distinction procrastinated significantly less than those who obtained the normal certificate. We use student IP addresses to geolocate them, and show that students from the United States behave differently and obtain better results than students from the rest of the world in both courses.

All these relationships are correlational; they do not establish that, for example, procrastination causes worse outcomes and that getting students to avoid procrastinating would improve their outcomes; perhaps students who procrastinate would not do as well even if they did not procrastinate. Therefore, we outline potential experiments to establish causal relationships between information provided students and their actions, and between their actions and achievement. We describe results from experiments conducted in Martinez (2014a, 2014b) that use email nudges to induce greater and more timely effort from students. These informational emails do not change the course architecture in any way, so they offer a very low-cost way to improve student outcomes.

Finally, we discuss the opportunities and challenges of survey data in MOOCs. Currently each course at Coursera asks its own questions and data are only shared at the university level. This means that students are asked about their age, gender, etc many times, – which, in turn, probably explains the very low response rate in MOOCs' surveys. We propose as a solution the creation of user profiles so that questions are only asked once. Finally, we think that some questions like, "Do you intend to complete this course?" should be asked as part of the enrollment process to learn more about student motives and their predictability of later actions.

The remainder of this paper is organized as follows. In the next section, we discuss related research. Section 1.3 describes the data. Section 1.4 explains the computational challenges. In section 1.5, we show how students who choose to drop out are different than students who choose to continue, and discuss how course design may play an important role in persistence. Section 1.6 discusses the role of the forums and show a positive correlation between forum participation and achievement. Section 1.7 explores the roll of video interaction in student achievement. Section 1.8 exposes how procrastination is negatively correlated with achievement. Section 1.9 discusses the value of certification and how it can affect student effort and achievement. Section 1.10 shows how we can link students across different courses. Section 1.11 explores the geolocation data provided by IP addresses, and shows how students from the US behave different than students from the rest of the world. Section 1.12 discusses the possibilities and limitations of survey data. Finally, we conclude in section 1.13.

1.2 Related Research

Einav and Levin (2013) discuss how Big Data will transform economic policy and research, business, government and other aspects of the economy. The role of economists and statisticians in this data-driven world has steadily increased with an ever blurring line between their traditional disciplinary backgrounds and those of computer scientists and systems engineers. Novel data types have opened the doors to cross-disciplinary work that allow new approaches to answering old questions while also eliciting new questions about individual behavior in an online environment, see Varian (2014).

1.2.1 Assessing Effort, Engagement, and Learning Outcomes

Most studies find that student effort, and not just ability, is positively correlated with academic success and achievement; see for example, Carbonaro (2005), Johnson, Crosnoe, and Elder Jr (2001), and Marks (2000). A limitation of these studies is that they rely on measures of effort that are reported by either the student or the teacher. A concern is that such appraisals are prone to inaccuracies or underlying biases. Self-assessment may be subject to substantial misrepresentation for a myriad of reasons. These include deliberate falsification, inaccurate record keeping, or other underlying student characteristics resulting in biased results. Teacher reports measures may fail to accurately assess time and effort, and may even confirm biases based on ex-post performance. Additionally, a teacher may not be able to uniformly evaluate each student which can result in certain effort strategies being over or under emphasized in effort measures.

Due to the online nature of MOOCs, much of the materials and interactions needed for class are available within the online course website. Therefore, usage can be monitored and measured. As a result, we are able to observe and quantify a greater percentage of actual student effort than in a more traditional classroom oriented structure. Furthermore, MOOCs allow for passive data collection on an incredibly large scale, a true benefit of the present Big Data age. Not only is it relatively cheap and easy to obtain an enormous amount of data, which itself is an advantage over earlier classroom data collection efforts, but also the data are collected in a non-intrusive and non-apparent manner. This limits our concern of potential Hawthorne effects.[2] As a result, we are better able to trust the data as legitimate behavioral measures.

When quantifying effort, effort itself must first be defined. Carbonaro (2005) separates effort into three different types: rule-oriented (compliance with the most basic rules and norms required by the school), procedural (meeting the specific demands set forth by a teacher in a particular class, including completing required assignments and participating in class discussions"), and intellectual (applying cognitive facilities towards understanding the intellectual challenges posed by the curriculum). We are not interested in rule oriented effort as we assume basic rule compliance of all students in order to complete the class. Using MOOCs, we have access to basic student information, such as whether or not a student completed a particular assignment or submitted responses for a quiz on time, thereby easily providing information to investigate procedural effort.

We do not focus on motivation for effort as is done some earlier research (Bong and Clark (1999) is one example). In this paper, we are not concerned as to why a student expended effort, as it is no easier to gauge motivation in a MOOC than

[2]The Hawthorne effect (also referred to as the observer effect) refers to a phenomenon whereby workers improve or modify an aspect of their behavior in response to the fact of change in their environment, see McCarney et al. (2007). For example, a request from the instructor to record how much time they are spending in the course can induce students to exert more effort.

in a classroom. Moreover, we do not make a distinction between engagement and effort as some researchers do (see as an example Newmann (1992)). Whether or not a student possesses the psychological component that Newmann alleges to be the distinguishing feature that separates engagement from effort is not relevant to our assessment of whether effort occurred. We restrict ourselves to what we can observe and quantify, which is much more extensive than in a traditional classroom. These measures capture components of what other researchers might separately distinguish as indicators of engagement or motivation.

Special attention is paid in our analysis below to the role of procrastination in academic performance. Previous research has considered many different measures for student procrastination: PASS (Solomon and Rothblum (1984); Owens and Newbegin (1997); B.L. Beck (2000)), Lay's procrastination scale (Lay (1986)), Tuckman's procrastination scale (Tuckman (1998); Michinov et al. (2011)), automated/digital assignment-related tracking measures (for example, the software *Homework Manager* as used in Rotenstein, Davis, and Tatum (2009)), and various other forms of teacher or student reported information (see for example, Owens and Newbegin (1997) or B.L. Beck (2000)). Though there are findings of mixed or no significant relationship between procrastination and measures of achievement (course grade (Solomon and Rothblum (1984)), final exam/test grade (Lay (1986); Tuckman (1998)), and others (B.L. Beck (2000))), the majority of findings indicate a negative relationship (across similar measures, for example see, Lloyd and Knutzen (1969); Semb, Glick, and Spencer (1977); Rothblum, Solomon, and Murakami (1986); Owens and Newbegin (1997); Van Eerde (2003); Howell et al. (2006); Rotenstein, Davis, and Tatum (2009); Michinov et al. (2011)). Furthermore, past research has found negative correlations within specific subject areas, for example mathematics (Akinsola, Tella, and

Tella (2007)) and writing (Fritzsche, Rapp Young, and Hickson (2003) and Ariely and Wertenbroch (2002)). Much of this research, however, relies on self-reported evaluations or measures of procrastination. As discussed above, passive observance of behavior, as done in this paper, as opposed to self-reported measures, are an unbiased and more accurate indicator of student activity. Where other researchers do collect actual student behavioral data, the focus is on one type of assignment or one classroom setting. (Rotenstein, Davis, and Tatum (2009)). We contribute to the literature by not only collecting actual student behavioral data, but by also studying assignments under different classroom settings.

In the same light, several researchers have noted the importance of social relationships for student learning and effort, especially in online learning environments (see for example, Carbonaro (2004); Garrison, Anderson, and Archer (2001, 2010)). Additionally, earlier research has suggested that voluntary student forum participation is correlated with academic performance (Cheng et al. (2011)). Whereas those papers set out to determine whether social relationships and forum interaction translate into academic success, we view forum participation only as a component of a larger effort and ignore the psychological aspect. Regarding the Cheng et al. (2011) research specifically, we seek to extend their results by further controlling for other measures of effort. Additionally, they explain failing to find a significant relationship between forum participation and assignment scores as "not surprising" since the nature of the assignments they studied was "not tightly related to the course materials...." We analyze data that, by design, is closely related to the presented course materials.

1.2.2 Prior Analysis of MOOCs

Although MOOCs are relatively new, the research interest on them is rapidly growing. Liyanagunawardena, Adams, and Williams (2013) found in a systematic study of literature published through 2012 that there had only been 45 distinct articles where MOOCs or their use were the primary focus. However, the pattern of publication shows a quickly increasing trend (one in 2008, one in 2009, seven in 2010, ten in 2011, and 26 in 2012). Of those, only 15 were classified under the broad heading of "Educational Theory." In 2013, research using MOOCS has continued, but it is evidently still in an early state. The nascent research is still trying to understand who enrolls in MOOCs and why (see, for example, Christensen et al. (2013)) and high level trends of completion and engagement (see, for example, Kizilcec, Piech, and Schneider (2013)). Those researchers who have made use of the MOOCs data have typically limited their focus to classification or broad associative observations. Kizilcec, Piech, and Schneider (2013), for example, present a simple classification method that identifies a small number of longitudinal engagement trajectories in MOOCs. Research is beginning to emerge at fine levels of specificity as evidenced by Breslow et al. (2013)), and this paper seeks to extend much of that work. In our research, we make use of the extensive classroom and limited survey data to uncover more subtle descriptive trends associated with student effort and performance, as well as to identify how survey data might be improved for future research.

1.3 Data

By way of demonstrating the richness of the information available from MOOCs, this study uses data from two Coursera courses. The first is *Foundations of Business Strategy* (FBS) taught by Professor Michael J. Lenox from the Darden Graduate School of Business Administration at the University of Virginia. This Coursera course enrolled $84,377$ students enrolled and ran from March 4, 2013 to April 14, 2013. The second is *The Modern World: Global History since 1760* (MWH) taught by Professor Philip Zelikow of the University of Virginia. This course had $46,575$ students enrolled and ran from January 14, 2013 to May 7, 2013. $7,253$ students were enrolled in both courses.

The data include time-stamped logs of student activities such as lecture views, submission of assignments, participation in forums (e.g., threads views, posts, and up-votes logs), clickstream logs (logs for tracking user activity on the course website), page views, lecture video interaction (e.g., video seek events), and geolocation information from Internet Protocol (IP) addresses.

1.3.1 Foundations of Business Strategy (FBS)

Foundations of Business Strategy is divided into weekly modules. Each weekly module consists of an introductory video, a reading from the strategist's toolkit, a series of video lectures, a quiz, and a case study to illustrate points in the lectures. Students wishing to receive a Statement of Accomplishment must satisfy the following criteria:[3]

[3] Extracted from the course wiki, available at http://goo.gl/yPb0du

1. Pass the weekly quizzes by scoring 90% or better on each individual quiz. Students have the opportunity to take each quiz as many times as they would like. The score of record is the best score on each quiz.

2. Submit a final project (a strategic analysis for an organization of the student's choosing), and score at least 50 points out of 100 points.

3. Assess five peers' strategic analysis using the peer assignment function.

Out of the $84,377$ students enrolled in this course, only $1,902$ obtained a Statement of Accomplishment. In Section 5, we discuss why we do not consider the $84,377$ students when thinking about dropout rates, but only those who took the first quiz.

1.3.2 The Modern World: Global History since 1760

The course wiki for The Modern World: Global History since 1760, available at goo.gl/po3mrf, states the following: "This course is a survey in modern world history for students, beginning or advanced, who wish to better understand how the world got to be the way it is today. Each week has a set of video presentations organized around a theme. There are five to ten of these relatively short presentations a week, each devoted to a particular topic within that week's theme. Video presentations are accompanied by quiz questions and optional reading assignments to reinforce your grasp on the factual material being covered and some of the interpretive problems. There are also weekly quizzes, based only on each week's video presentations; these quizzes are the only graded component of the course."

Course grades are based solely on weekly quiz performance (20 questions/week); the in-video quiz questions are not scored. The grading formula is as follows: Total of 280 questions (20 x 14 weeks) scored at 0.4 apiece with all scores over 100 counted as 100. Statements of Accomplishment are available to students according to two tracks:

1. Distinguished - Students must achieve a cumulative quiz average of 85% or better.

2. Passing - Students must achieve a cumulative quiz average of 65% or better.

Out of the 46,575 students enrolled in this course, 4,939 obtained a statement of accomplishment, and 3,974 of these were distinguished.

1.4 Computational Challenges

Coursera data are very rich. For example, we are able to see the exact time each student accesses a forum thread, clicks play on a given video, pauses it, changes the speed, and every answer given in every attempt for each quiz. It is simply not feasible to have all that data in a rectangular array for easy access by conventional statistical software applications.

Coursera uses two mechanisms of data collection across all of their courses. For each course, the researcher receives a MySQL data dump. This database includes:

- Versioned copies of instructions for all surveys and assessments, including quizzes, homeworks, exams, in-video quizzes, assignments, and peer-graded assessments.

- Time-stamped and versioned copies of student responses for all surveys and assessments.

- Time-stamped logs of student activities such as lecture watching, assignment submission, and forum behavior.

- All forum content, including upvotes logs and the full text of posts and comments.

- Student registration information.

In addition, a json file contains the clickstream logs. This file contains information about every click a student makes in the course website. For example, we are able to see the exact time each student clicks the play button in the media player or visits any given forum thread.

Although these rich data about student engagement and achievement present great opportunities for research, they also create challenges. For example, the json file for MWH is a 16GB file with hundreds of millions of strings of text that need to be processed in order to generate useful tables. Once these tables are generated, the data have to be merged with data from other tables. This task is simply not feasible for a single computer. Parallelizing these operations allows us to reduce the amount of time to do the overall analysis from days to hours.[4]

In addition to the Big Data challenge related to computer cycles, we also face a storage challenge. To work with these data, it is necessary to obtain the Institutional Review Board (IRB) approval. Because the data contains student unit records, storing it in a public cloud is not an option. The data has to be stored on a secure MySQL.[5]

[4] Almasi and Gottlieb (1989) defines parallel computing as a form of computation in which many calculations are carried out simultaneously, operating on the principle that large problems can often be divided into smaller ones, which are then solved concurrently.

[5] The DBM2 system, which can run MySQL and can be accessed by the ITC cluster, is not suitable for large data – the default limit is 100 MB. We were not able to get permission to run a MySQL server on the ITS cluster. This challenge was resolved by UVACSE lending some of its disk

A known issue with MySQL is that it keeps a log of all transactions made with the database. This log quickly exceeded the disk space partitioned for applications, which caused the computer to crash. This challenge was resolved by the University of Virginia Alliance for Computational Science & Engineering (UVACSE) re-installing the database driver so that the data and log could be stored on a larger partition of the disk.

An anticipated challenge is that the current software may not be able to handle larger data. For example, loading the data provided by Coursera into the database is a sequential process. When files from multiple courses are received, loading the data in a timely manner may be problematic. Similarly, with parallel R, there is a limit (2 GB) to the amount of data collected from parallel processes. So, some analysis may need to be done sequentially to avoid a software failure.[6]

1.5 Dropping Out

Before thinking about dropout rates, it is important to understand what enrollment means in a MOOC. In order to be enrolled in a traditional classroom a student, or parent, has to pay some cost and meet some requirements. For example, in order to enroll in Economics 2010 at the University of Virginia (UVa), a student needs to be admitted into the university and pay tuition. The selection process is formidable: admitted students typically have high SAT scores and high school GPAs.[7] To enroll

space to the project and providing a dedicated system that has MySQL running on it. The node with the storage space was mounted to the dedicated system so that the data could be accessed by the database.

[6]We are investigating software that will be better suited to work with data of larger magnitudes. Jacalyn Huband, from the UVACSE, is looking into NoSQL databases and commercial R packages (by Revolutionary Analytics) that are designed specifically for big data.

[7]See http://www.parchment.com/c/college/college-1583-University-of-Virginia.html

in any MOOC, a student only needs to click on a link; there is no need to have any degree, go through any admission process or pay anything. Hoxby (2014) argues that high dropout rates in MOOCs may simply reflect trial and error since the cost of signing up is very small. The MOOC system is more similar to the public university system in some U.S. states and countries than to UVa. For example, in Argentina any person with a high school degree can enroll in a public university without an admission test and without having to pay tuition. Official statistics show that out of 100 students, 23 get a degree in a public or private university. The most inefficient university produces only 3 graduates out of 100 students.[8] When thinking about dropout rates in MOOCs, one should be thinking about free, open-admissions Argentinian universities, not selective universities in the United States. However, unlike brick-and-mortar courses where an additional student attending limits capacity, MOOCs distinguishing feature is that they have zero marginal cost.

1.5.1 Foundations of Business Strategy

FBS required students to score a 9 on each quiz in order to obtain a statement of accomplishment, but it also allowed them to retake the quizzes as many times as they wanted and had a single deadline the last day of the course for all quizzes. All these course design choices probably affected students achievement, but it is not possible to disentangle how. To start addressing this question, in this section we show that students who drop out after Quiz 1 made different choices than those who continue. We show that students who drop out interact with videos and forums less than students who continue, and procrastinate more. We finally propose experiments

[8] See “Necesitamos más graduados” http://goo.gl/Q98vj4

that would help disentangle whether there is a causal effect and ultimately improve student achievement.

Table 1.6 shows the number of students who took Quiz “X” and “Y” for FBS. For example, $11,183$ users took Quiz 1 but only $2,677$ of those also took Quiz 6. Notice that out of the $84,377$ students enrolled in this class, only 15% took Quiz 1. Therefore, when thinking about the dropout rate for FBS, we consider not the $84,377$ students who clicked the enroll button but rather the $11,183$ who took Quiz 1. If we do this, the dropout rate for this course is 82.99%. Panel (a) in Figure 1.2 shows the distribution of the number of quizzes taken by the students, conditional on taking any. $2,529$ students attempted all the quizzes, and still less, $1,902$, received a statement of accomplishment. This appears to reflect the low value of the certificate compared with the cost of writing a final project and evaluating the final project of at least 5 peers.

1.6 Forums

In a traditional classroom, students can go to office hours or talk to classmates if they are having troubles with a concept; the forums play that role in MOOCs. The forums are not only the place where students ask questions about the material and quizzes, but also a place to socialize with their peers. For example, typical forum threads will have subjects like “Quiz 1: Q1?” and “Spanish speakers study group”. This allows professors, and researchers, to observe which topics generate the most discussion, a difficult feature to observe in a traditional classroom. Moreover, forums may be a good tool to use in a bricks-and-mortar classroom.

In the FBS forums, $2,679$ threads were created. Table 1.1 shows the differences between forum readers and non-readers. Students who decided to read at least one forum thread before attempting Quiz 1 for the first time scored 0.14 points less in their first attempt, 0.26 points less in their best attempt, and have a drop-out rate 7 percentage points higher than students who decided not to read the forums. To understand this we need to think about selection into reading the forums. Students who have no trouble with the material do not need to go to office hours, or, in the MOOCs world, read forums. But the interesting question is whether reading the forums helps struggling students improve their grades and persist in the course. In order to answer this we would need an exogenous source of variation that motivates some students to participate in the forums. For example, the platform could randomly assign students who miss a question in a quiz to receive a notification that encourages them to check a forum thread that discuss that problem.

Examining forum participation and quiz retakes shows some evidence that forums may help students improve outcomes. Students who took Quiz 1 more than once and read the forums between their first and best attempt performed worse in both attempts than those who did not read the forums. This is likely explained by selection into reading the forums. Students who are struggling are more likely to spend more time in the forums than students who are not. It is important to note that these active forum readers improved more between their first and best attempt than the non-readers. Therefore, forum participation is positively correlated with achievement. In addition, we find that the dropout rate is lower for those students than for students that do not use the forums between their first and best attempt.

1.7 Videos

Video lectures lie at the heart of MOOCs. Table 1.2 shows the effect on achievement of watching the lectures before attempting the quiz for the first time and between the first and best attempt. Students who choose not to watch the videos before attempting the quiz for the first time scored, on average, 0.89 points less than those who watched the videos. Additionally, they scored 0.51 points less in their best attempt. Students who decided to watch videos between their first and best attempt improved 0.56 points more than those who did not watch any videos. Finally, the dropout rate for watchers is 13% lower than for those who did not watch any video between their first and best attempt. Once again, an experiment that encourages students to revisit the videos before attempting the quiz again could help us determine if this is a casual effect or if lecture-watchers have some unobserved characteristic that makes them less likely to dropout.

1.8 Procrastination

Thomas Jefferson, founder of UVa, famously said, "Never put off till tomorrow what you can do today."[9] In the previous sections we have shown the negative correlation between achievement and procrastination. Observational data shows a strong correlation between procrastination and negative quiz outcomes; this relationship would be very difficult to observe in a bricks-and-mortar classroom. Panel (a) in Figure 1.3 shows the correlation between when students choose to attempt Quiz 1 for the first time for FBS and the grade they obtain in their best attempt. We not only

[9] `http://www.monticello.org/site/jefferson/canons-conduct`

see a clear negative correlation between procrastination and achievement, but also much lower achievement for those student who attempted Quiz 1 for the first time after the second quiz is published. This is the case for all quizzes in FBS.

Panel (b) in Figure 1.3 shows the same negative correlation for MWH, but the lower achievement is even more pronounced after the second quiz is published. The reason for this is that students in MWH were penalized by 5% per day they were late in their assignments.[10] As in FBS this correlation is consistent across all the quizzes.

Figure 1.4 shows the distribution of minutes elapsed between students' first and last attempt for FBS Quiz 1. Half of the students spend less than 12 minutes between their first and best attempt, and 75% spend less than 42 minutes.

1.9 The value of the certificates

Economists and other social scientists have a long-standing interest in understanding the extent to which students and employees value degrees or "certification." At the core of the issue is whether certification holds value because it resolves an asymmetric information problem whereby it may be difficult for firms to infer true achievement and skills, see Spence (1973), Hungerford and Solon (1987), and Clark and Martorell (2014). MHW has two levels of certificates, with and without distinction. Table 1.3 shows students that obtained the certificate with distinction procrastinated, on average, 86.55 hours less between the moment the first quiz was published and when they took it than those who received the "normal" certificate. A randomized control trial could allow researchers to understand how students change behavior when better quality certificates are available and how that change in behavior ultimately affects

[10]Figure Panel (b) in Figure 1.3 shows raw scores without the 5% per day late penalty.

achievement.[11] Signaling models predict that changes should occur for specific groups of students, those for whom learning/performance is easy.

1.10 Linking students

A total of 123,699 were enrolled in FBS or MWH, and 7,253 of these students were enrolled in both courses. The ability to link students across courses allows researchers, for example, to study how an intervention in one course affects behavior and outcomes in another.

Table 1.4 shows how students who were enrolled in both courses differ from students enrolled only in FBS. For Quiz 1, students who were also enrolled in MWH performed 0.29 points better in their first attempt, but there are no statistically significant differences in their best grade nor in the fraction that obtained the certificate. If we restrict the sample to students who took Quiz 6, there are no statistically significant differences in performances but the proportion of students who dropout is 9 percentage points higher than for students taking MWH.

Having the ability to link students across courses can be important for future experiments. For example, how does a student change his behavior in course "B" when he is nudged to exert more effort in course "A". Moreover, we think that having the ability to follow students across courses will eventually allow us to create personalized nudges for different type of students.

[11] Coursera also offers the "Signature Track" service to verify identity which may, in turn increase student effort and achievement.

1.11 Geolocation

Every time a student logs into Coursera, their IP address is saved. Using the last available IP address for each user, we are able to geolocate them.[12] This information can be linked with other Coursera information like achievement, video behavior, etc. It is also possible to link this information to geographic information like the unemployment rate, average internet speed, local language, weather, etc.

Panel (a) of Figure 1.5 shows how students who took FBS Quiz 1 are distributed across countries. Out of the $11,183$ students that took Quiz 1, 2,278 live in the United States, 818 in India, and 398 in Brazil, and smaller numbers live in a range of other countries. Panel (b) shows the distribution for students across the United States: most of these students live in California (355), followed by New York (175), Illinois (146), and Virginia (132). Figure 1.6 restricts these distributions to students who obtained a statement of accomplishment.[13] Out of the $1,902$ students who obtained a statement of accomplishment, 240 live in the United States, 133 in India, and 75 in Spain. Out of the 240 who live in the United States, 38 live in California, 21 in Illinois, 18 in New York, 15 in Florida, 14 in New Jersey, and 13 in Virginia.

Table 1.5 shows differences in outcomes and choices for Quiz 1 for students from the United States and students from the rest of the world. On average, students from the United States taking FBS scored 0.42 points higher on their first attempt and 0.15 points higher on their best attempt. They also attempted the quiz 0.1 fewer times and procrastinated 45.16 fewer hours before attempting the quiz for the first time. U.S. students taking MWH scored, on average, 0.73 points higher, and procrastinated

[12] For IP addresses within the US we are able to determine a student zip code from his IP address information.
[13] For MWH see Figure 1.7.

66.36 fewer hours than students from the rest of the world. It is likely the effectiveness of a nudge will vary by country because of cultural differences.

Figure 1.8 shows the grade distribution for the first attempt at Quiz 1 in FBS. We can see that only 17.78% of students obtained the required 9 or better needed to qualify for the statement of accomplishment. Figure 1.9 shows the grade distribution for the best attempt at Quiz 1. In this case, 66.09% scored 9 or 10 points. Finally, Figure 1.10 shows the distribution of number of attempts for this quiz. On average students attempted the quiz 2.5 times. This shows that scoring a 9 requires a non-trivial amount of effort. If certificates are not very valuable, students may choose to drop out because the difficulty of the course is too high, which could explain the low completion rate. To address this, some courses like MWH offer two certificates instead of just one: the first with lower requirements, and a second with higher requirements and a distinction.

Table 1.7 shows the differences between students who decided not take Quiz 2 after taking Quiz 1, "Drop-out," and those who decided to take it, "Continue." The first row, "First Grade," shows that on average, students who drop out scored 0.56 points less on their first attempt to Quiz 1 than those who decided to continue. The second row, "Max Grade," shows that on their best attempt students who drop out scored, on average, 1.37 points less than those who decided to continue. Moreover, these dropouts have an average score lower than 9 which means that they would not be able to obtain a certificate for this course. One possible explanation for the high dropout rate is the very strict policy of requiring a 9, on each quiz in order to obtain a certificate. The next rows examine students' choices. "Number of attempts," shows that dropouts attempted Quiz 1, on average, 0.66 fewer times than students who decided to take Quiz 2. "Play 1," shows that on average students who drop out press the play button before attempting the quiz for the first time 3.91 more times than those who take Quiz 2. In the next row, "Cond. Play1," we condition this on students who press the play button at least once. For this population, dropouts press the play button 4.47 fewer times. This is probably explained by the fact that some students already know the material and are taking the quizzes without watching all the videos.[14] Similar patterns are found for "PlayM", which records patterns between the student's first and best attempt on Quiz 1.

Because successful students watch the videos more often and for longer than students who ultimately drop out, an important question is whether or not it is possible to decrease dropout rates by encouraging students to watch the videos. Experiments to test this could nudge students by simply sending them an email, or the platform could require a certain level of video interaction before allowing students to take the quiz, or the quizzes could be embedded in the videos. In the next rows, we examine how forum behavior is correlated with dropout decisions. "TotalViews1" shows that students who continue to the next quiz read, on average, 1.21 more threads before attempting Quiz 1 than students who dropout. After conditioning on reading at least one, "Cond. TotalViews1," this difference increases to 2.02. Once again, an experiment that encourages forum participation would allow us to determine whether there is a causal relationship.

Finally, we look at the number of hours between the posting of the quiz and the student's first attempt on Quiz 1. We find that dropouts procrastinate, on average, almost 16 hours more than those who decide to continue on to Quiz 2. Determining whether or not this is a causal effect is very important for course design. For

[14]Note that ideally we would like to know for each video how long a student spend watching it instead of just counting clicks. This information is currently not collected by Coursera. We have made that feature request and are also looking at other platforms that collect these data.

example, some courses have strict deadlines for quizzes, others soft deadlines with penalties for late days, while others no deadlines at all. If we are able to determine that procrastination causes bad performance and dropouts, then the optimal strategy is to discourage it. For example, MWH discourages procrastination by penalizing studentgrades by 5% for each day they submit their quiz late. But, it is possible that some other unobserved characteristics cause this correlation between delaying a quiz attempt and poor performance on the quiz, then, the optimal strategy is not to have deadlines at all.

1.11.1 The Modern World: Global History since 1760

In contrast to FBS, in Professor Zelikow's course students are allowed to attempt the quiz only once and they are penalized by 5% per day they are past the deadline. On the other hand, not doing well on a single quiz does not disqualify the student from receiving the statement of accomplishment. Finally, a statement of accomplishment with distinction was also available.

Figure 1.8 shows the number of students that took Quiz "X" and "Y." For example, $13,479$ students took Quiz 1 but only $6,985$ of those also took Quiz 14. Panel (b) in Figure 1.2 shows the distribution of the number of quizzes taken by the students, conditional on taking any.

Table 1.9 highlights the differences between people who decided to not take Quiz 2 after taking Quiz 1, and those who decided to take it. Unlike FBS, in this course students can only attempt a quiz once, and they are penalized by 5% per day they are past the deadline. The first row, "Grade," shows the raw grade that students obtained in Quiz 1. Students who decide to drop out scored, on average, 2.37 points less than those who take Quiz 2. The second and third rows, "Play" and "Cond. Play", show that students who decide not to take the second quiz also press the play button fewer times before attempting on Quiz 1 for the first time. Similarly, the fourth and fifth rows, "TotalViews" and "Cond. TotalViews", show that dropouts read fewer forum threads before attempting the Quiz for the first time. An experiment that nudges students into interacting more with videos or forums would answer whether or not there is a causal relationship. Finally, dropouts procrastinate, on average, 178.76 hours more than students who decide to continue. It is important to remark that the 5% penalty has two effects when comparing the two courses. First, students that continue in the course procrastinate, on average, 46.08 hours less. Second, students that dropout procrastinate, on average, 116.87 hours more than in FBS.

1.12 Survey data

Surveys are a powerful tool for collecting data, but Coursera's survey system is a weak point. For example, their system does not allow for skip patterns for more efficient responses, nor personalized questions by geolocation. A major problem arises because students who sign up for multiple courses are asked some variation of the same questions over and over again (e.g., what is you age, gender, education, etc) in each course. If Coursera had a central repository where researchers can choose the questions they need to ask and the survey should be personalized by the student in order not to ask the same question again and again. But, the biggest challenge of using survey data for research is the incredibly low response rates. Paying students to complete a survey is not a reasonable option for MOOCs. But, instructors could require completing the survey as part of the requirement for getting a certificate.[15] This is not all that different than what Facebook or Google do in exchange for a free service.

Out of the 13,923 students who took MWH Quiz 1, only 4,726 completed the survey.[16] Keep in mind that because of selection, and the low response rate, we cannot make any definite conclusion with these data. The purpose of this section is to show the potential for research using these rich data if the response rate were higher.

Figure 1.11 show the distribution of ages of MWH survey respondents. Table 1.10 shows that students under the age of 25 scored, on average, 0.25 points less than students age 25 and older. Young students also procrastinate, on average, 27.09 more hours before attempting the quiz. Finally, the proportion of young students that do not receive a certificate is 8 percentage points higher than for students age 25 and older.

Table 1.11 compares native English speakers to non-native speakers. Native speakers scored, on average, 0.76 points more than non-native speakers. Non-native speakers procrastinated, on average, 41 more hours than native speakers. Finally, the dropout rate is 3 percentage points lower for native English speakers. Taking into account selection into answering the survey, we expect the real difference in dropout rate to be actually much higher. In order to address this problem MOOCs can, for example, provide subtitles. For example, it would be easy to design an experiment to test how effective computer-generated subtitles are compared with human translation.

[15] The answer, "I prefer not to answer this question" should be an option for all questions, but it should not be the default as it is now.

[16] 282 completed the survey but not Quiz 1.

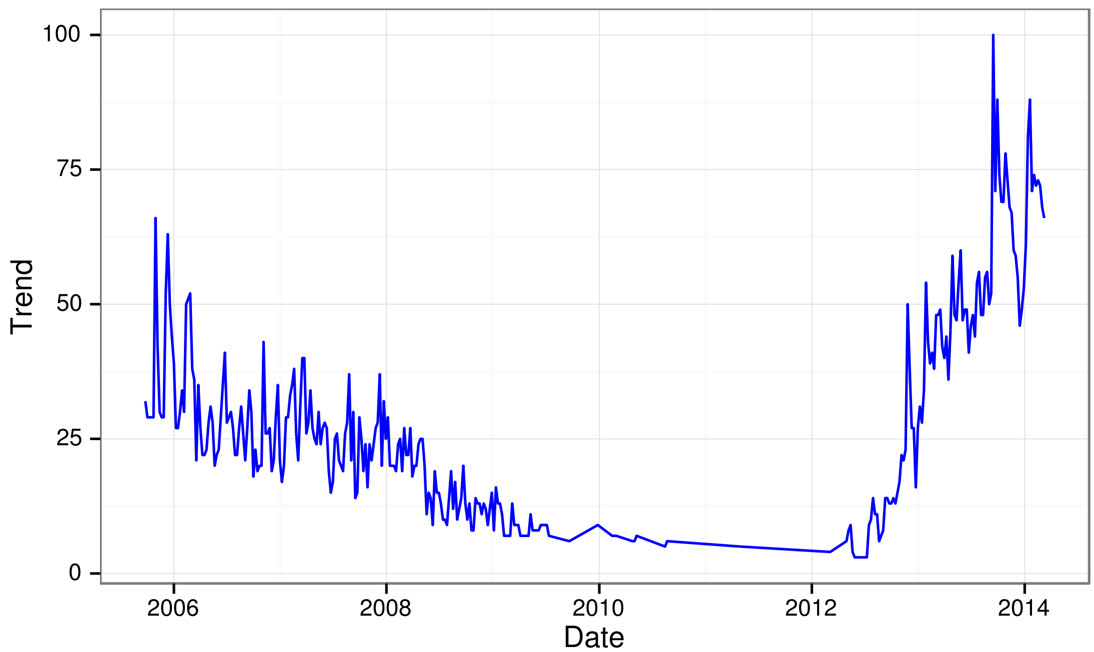
1.13 Conclusions

MOOCs generate a tremendous amount of data that can be used not only to improve MOOCs but also traditional classrooms. MOOCs offer an excellent opportunity to do new research on procrastination. We show that procrastination, as measured by delays in taking quizzes after they are posted, is negatively correlated with achievement on quizzes, but we need to determine if this is a causal relationship. Understanding this can not only affect MOOC design, but it can also affect the design of traditional courses. Fortunately, MOOCs are an ideal laboratory for testing this with high precision and very low cost. Experiments to reduce procrastination can involve either establishing or altering deadlines or sending students almost costless emails that give them information about the harm done by procrastinating.

Experiments to determine how students change their behavior when they are offered certificates of different values are also needed. This will allow us to determine how these changes in behavior affect outcomes.

Although Coursera collects very rich data, further improvement to their platform could greatly benefit research. First, we know from studying clicking behavior that collecting data on time use is very important to further our understanding of how achievement is produced, and ultimately help student succeed. Second, Coursera can take steps to improve their survey system and improve data sharing. A student who is taking 10 courses should not be required to answer 10 times how old he is; that data should be available to the researcher and only asked once. Coursera should create profiles for its users in the same way social network companies like Facebook do. We are convinced that MOOCs not only offer students a new way to learn but also offer researchers a new laboratory for understanding learning.

Figure 1.1: Google search interest over time for the word “MOOCs”



Note: This plot is generate using data from Google trends. Numbers represent search interest relative to the highest point on the chart.

Figure 1.2: Numbers of Quizzes taken, conditional on taking one or more

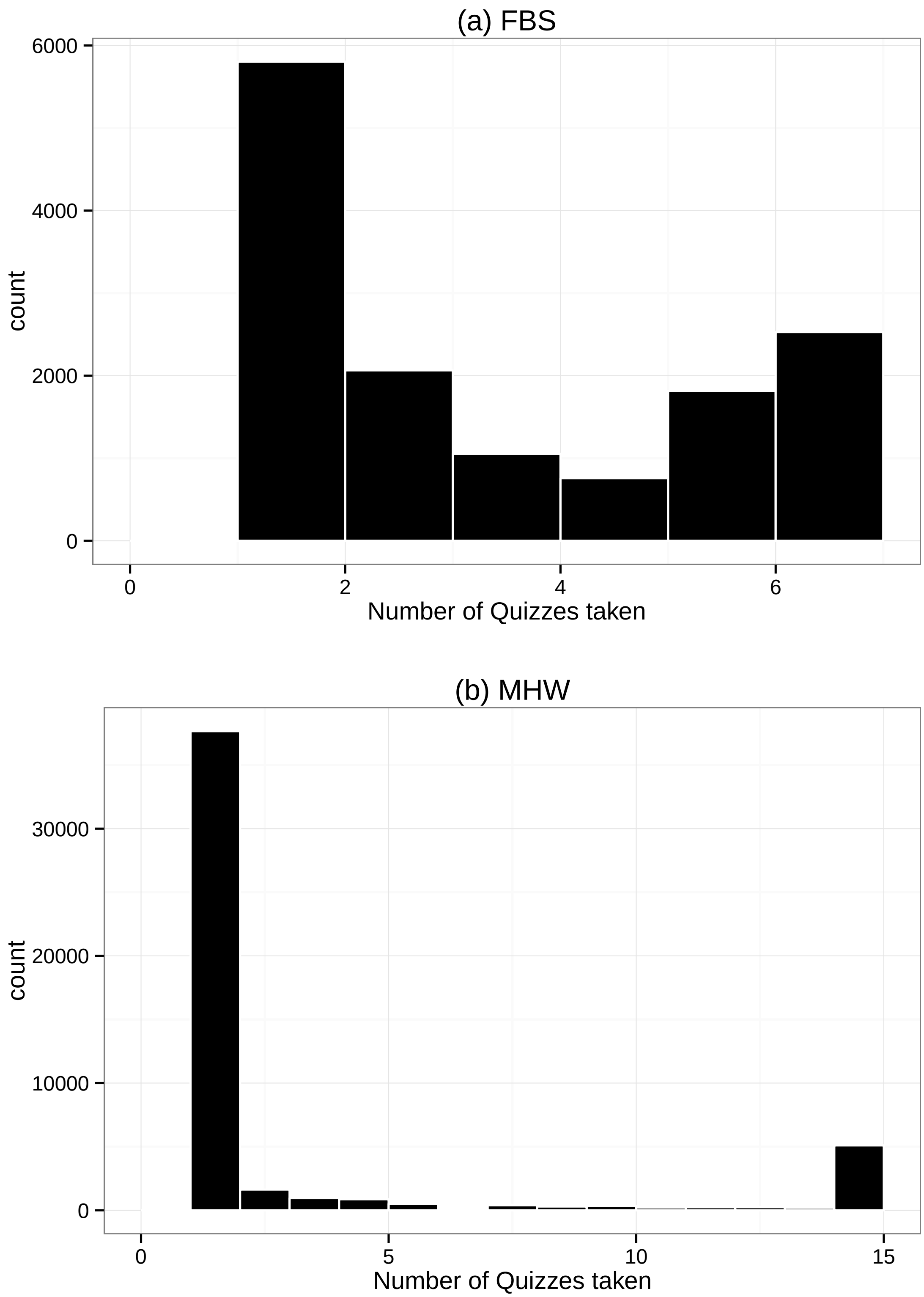


Figure 1.3: Date of the first attempt at Quiz 1 and maximum grade

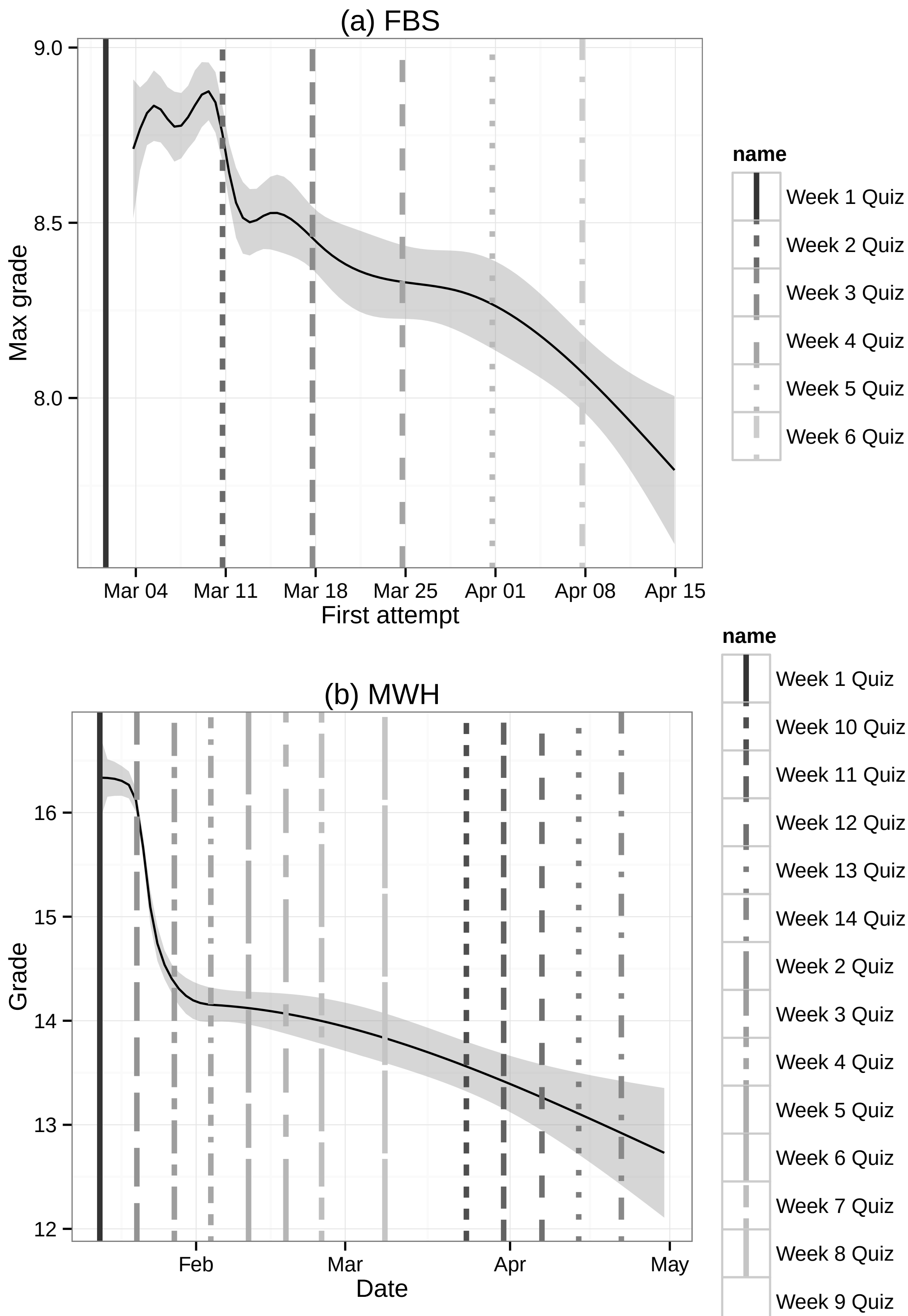


Figure 1.4: Distribution of minutes elapsed between first and last attempt for Quiz 1 (FBS)

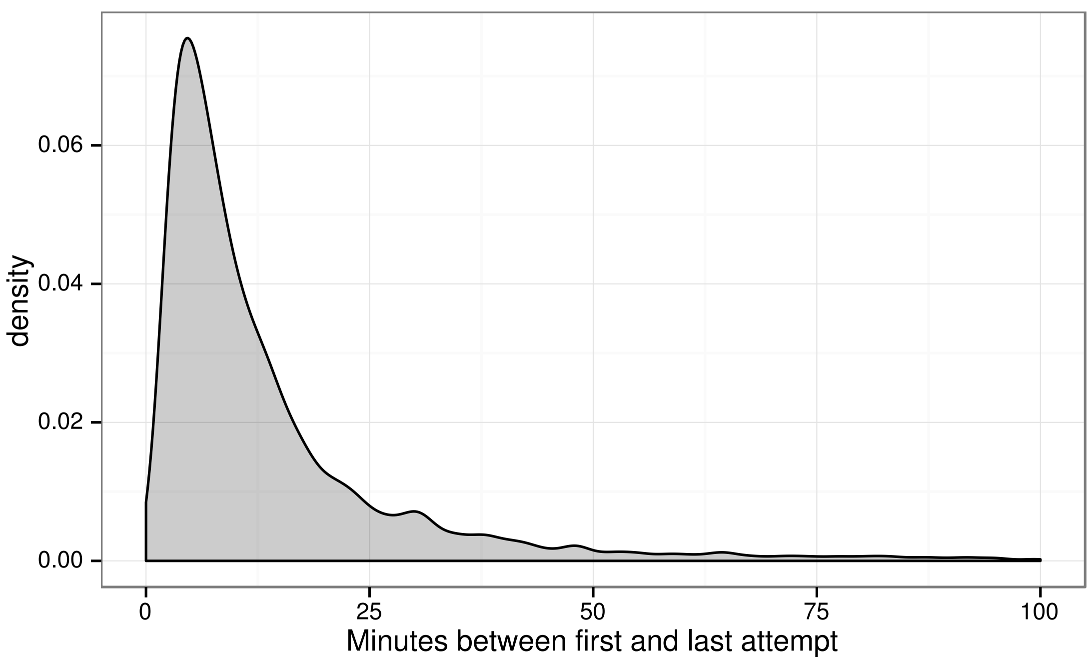


Figure 1.5: Geolocation of students whom attempted Quiz 1 (FBS)

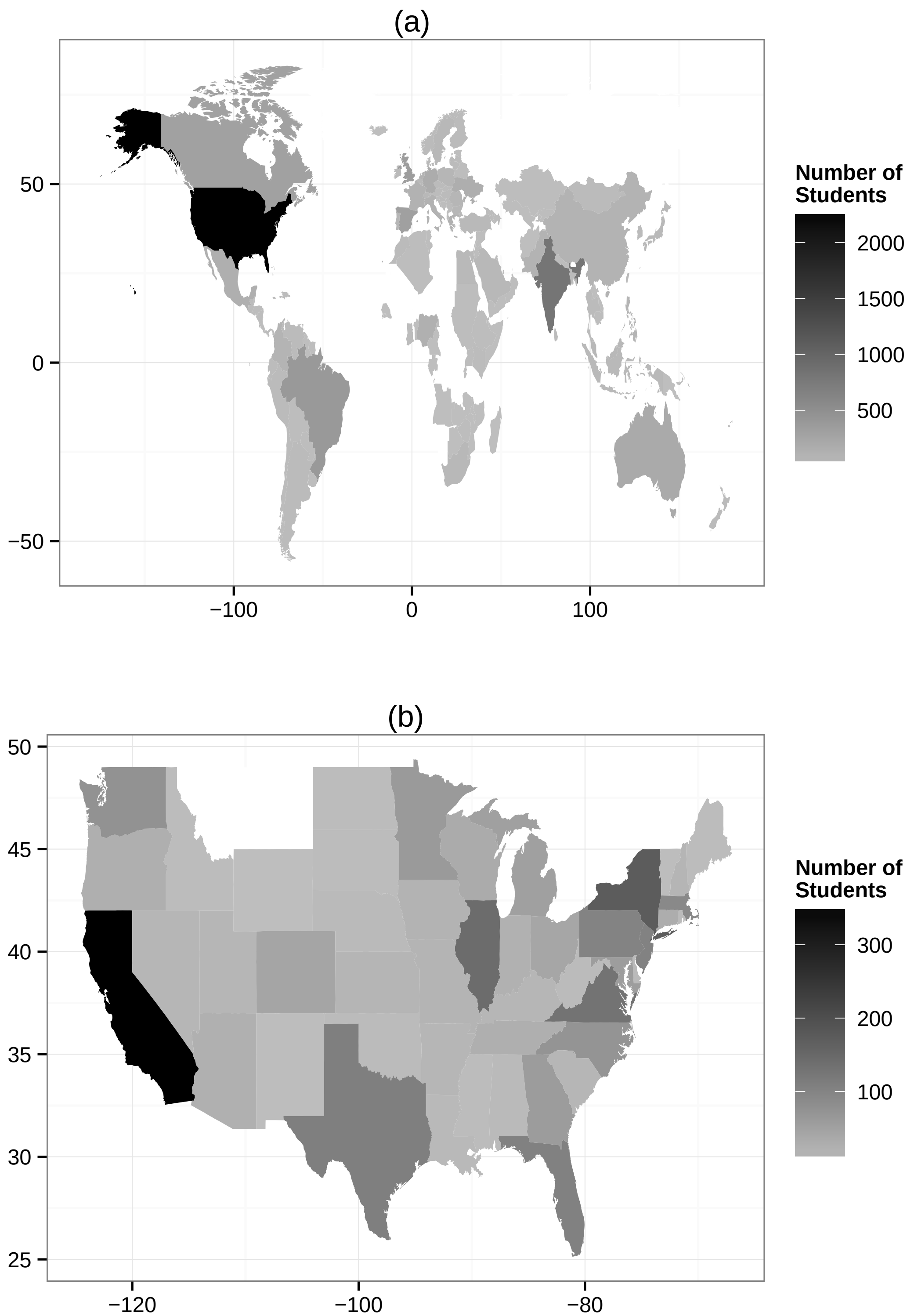
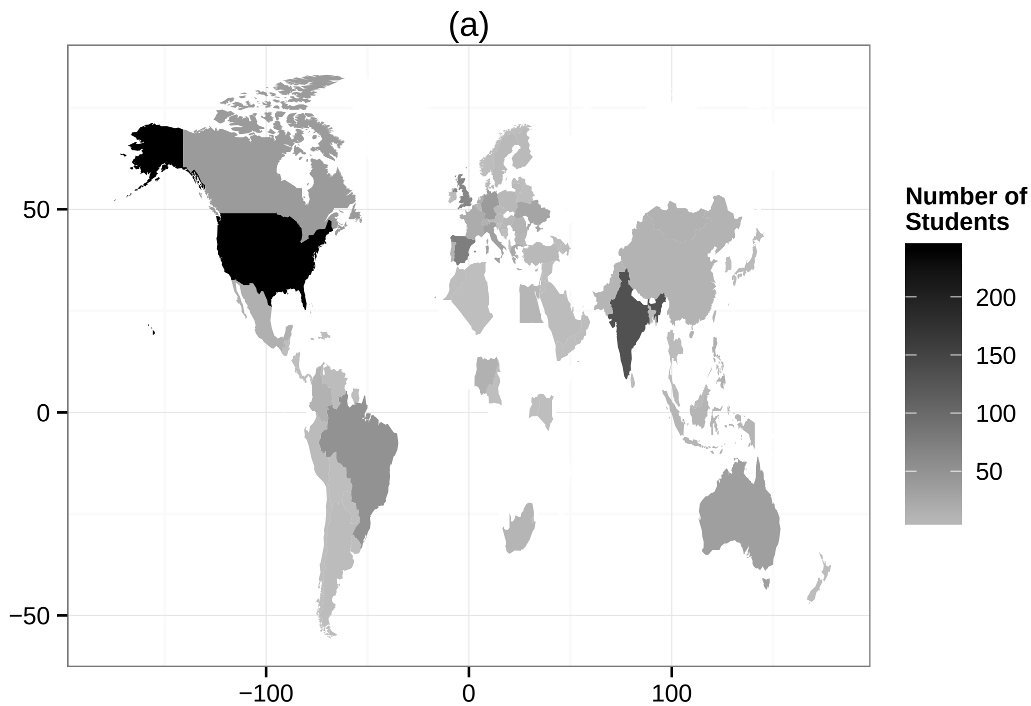


Figure 1.6: Geolocation of students who obtained a statement of accomplishment (FBS)



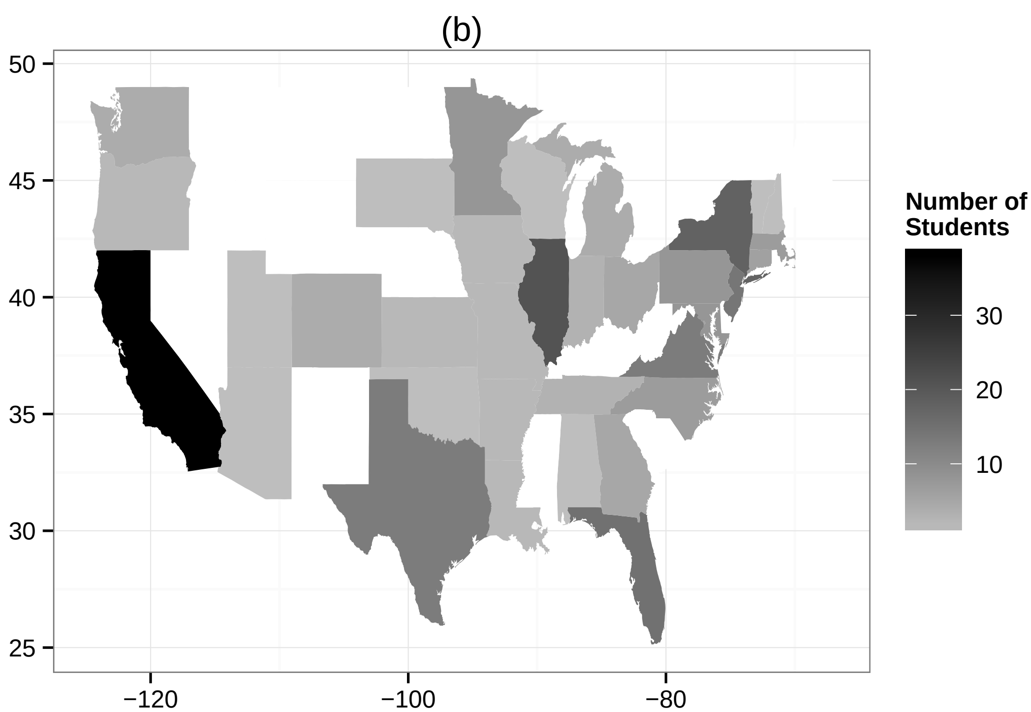


Figure 1.7: Geolocation of students whom attempted Quiz 1 (MWH)

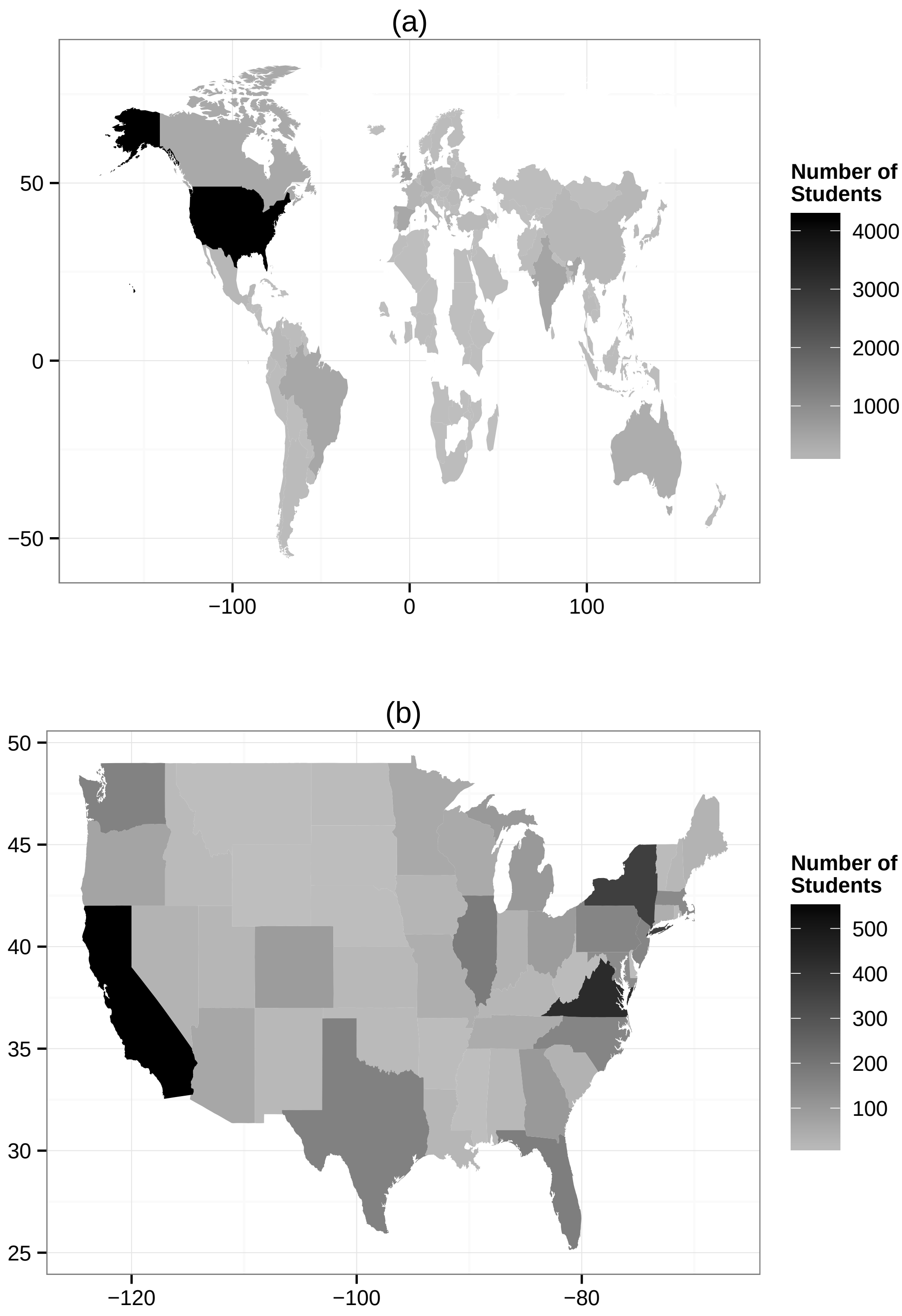


Figure 1.8: Grade Distribution First Attempt Quiz 1 (FBS)

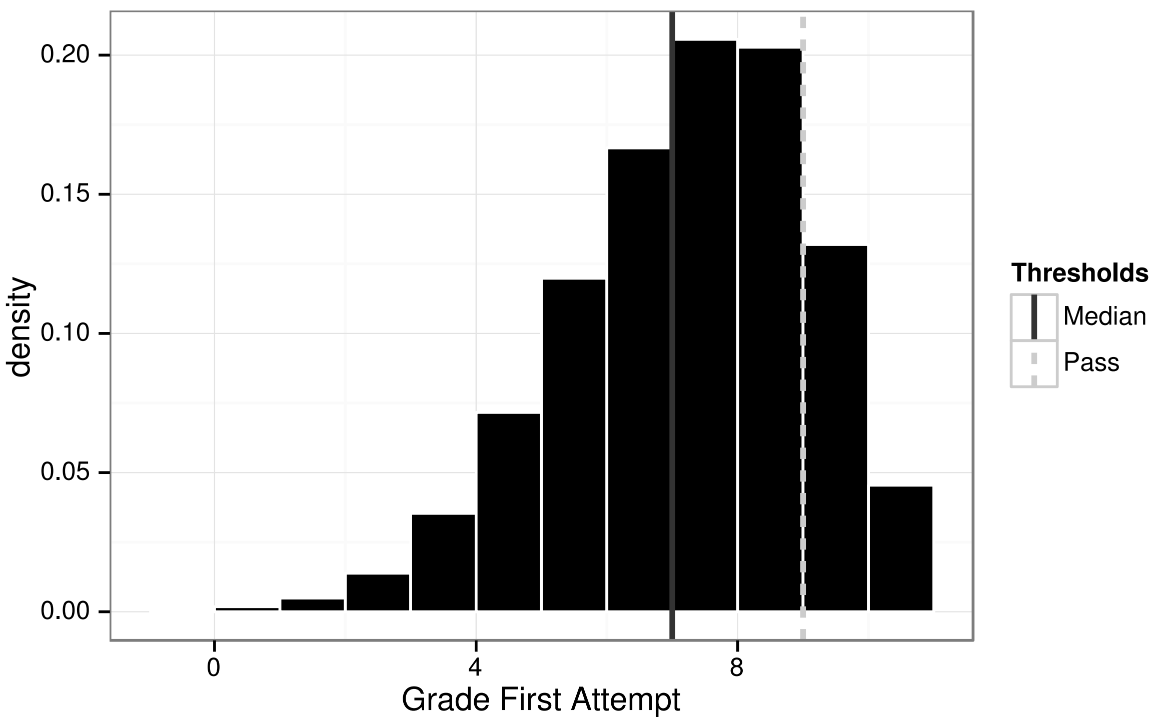


Figure 1.9: Grade Distribution Best Attempt Quiz 1 (FBS)

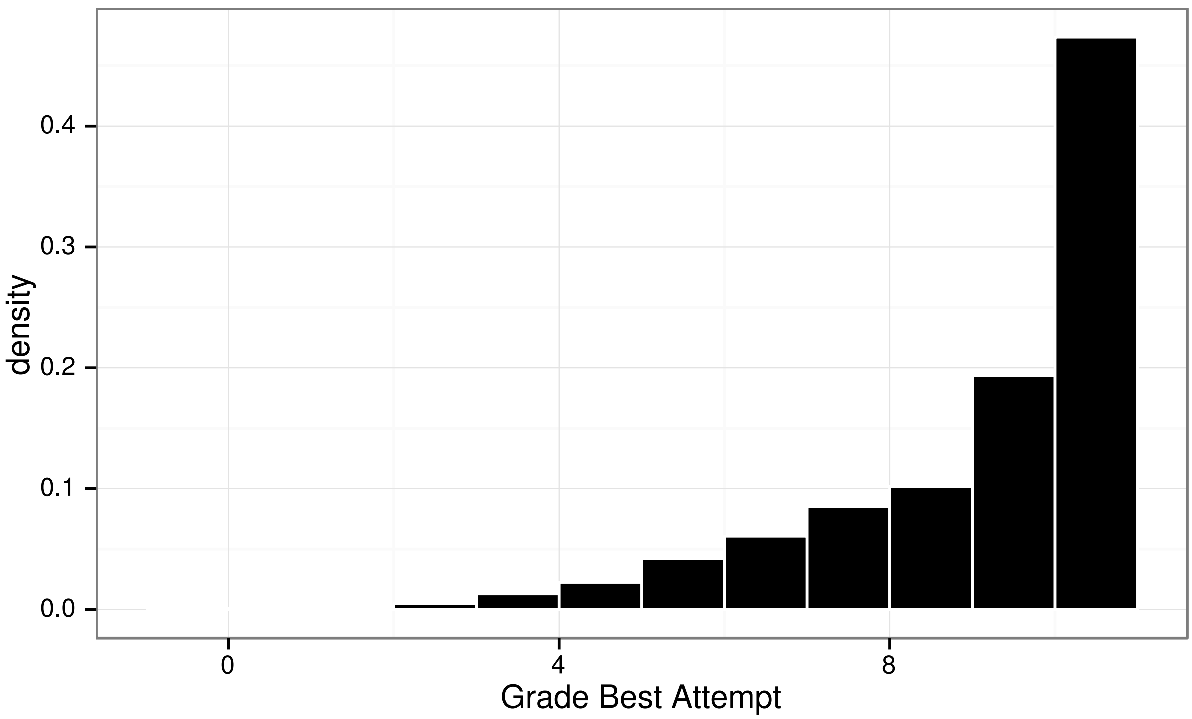


Figure 1.10: Number of attempts for Quiz 1 (FBS)

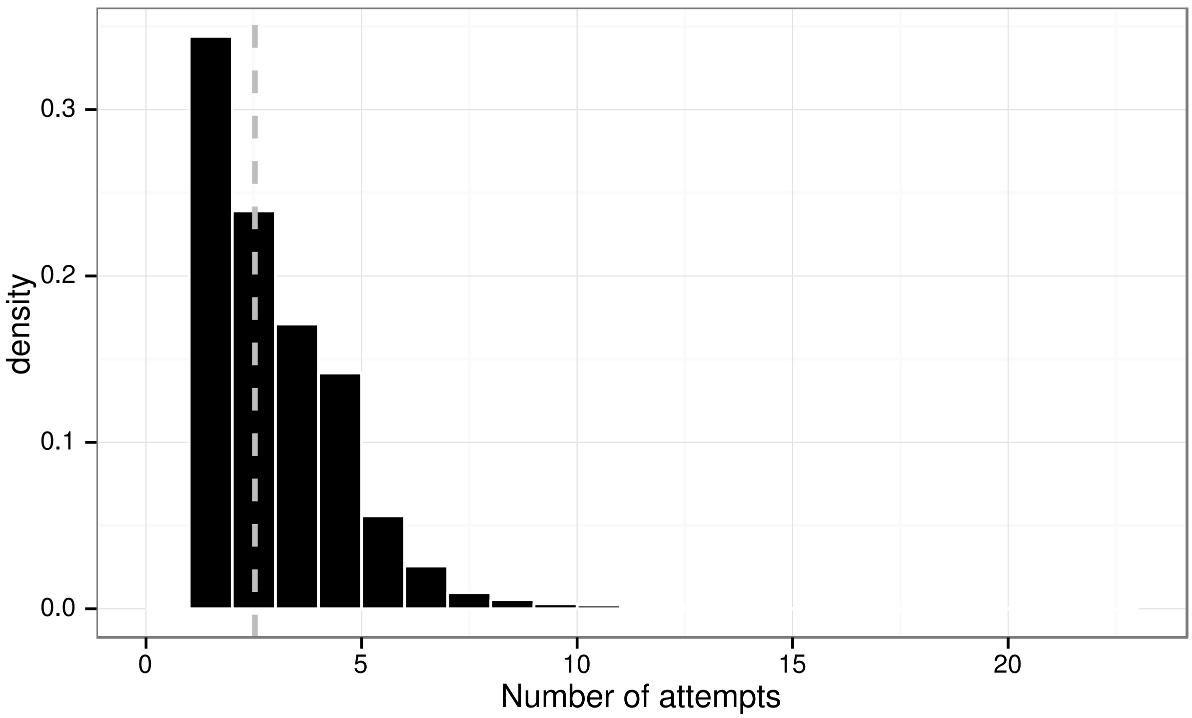


Figure 1.11: What is your age? (MWH)

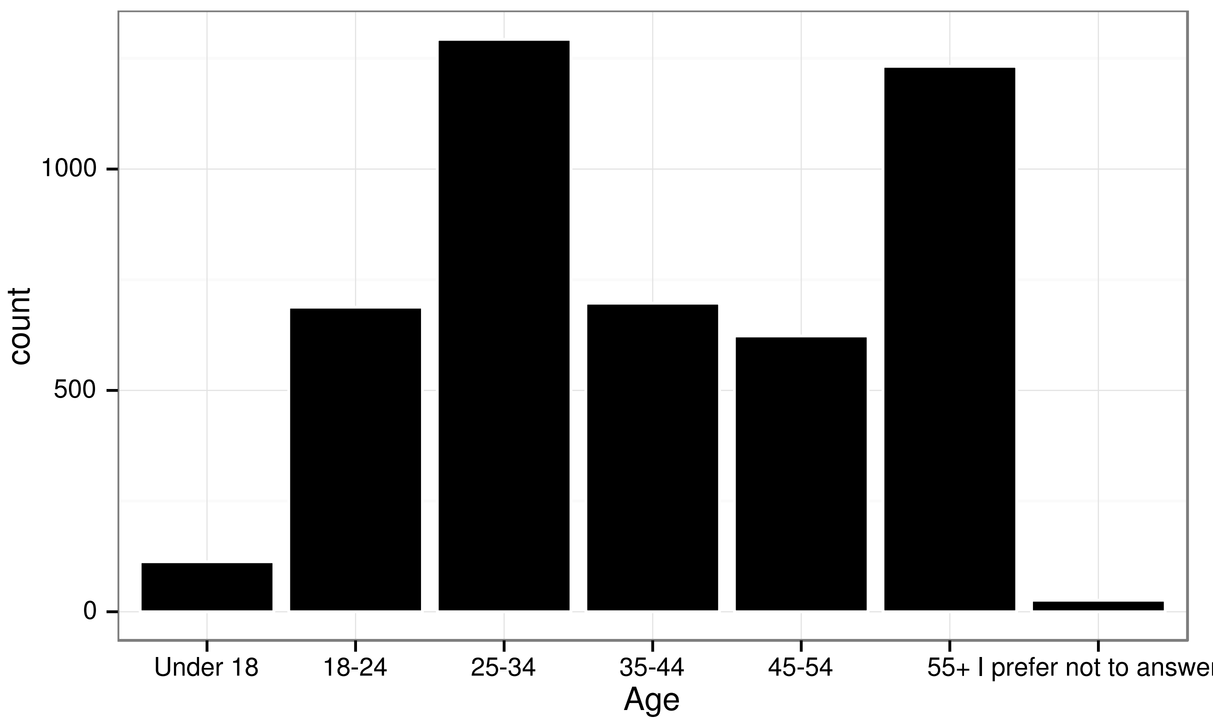


Table 1.1: Reading before the first attempt

	Readers	Non-Readers	p value
		Before First Attempt	
First Grade	6.92	7.06	0.00
Max Grade	8.87	9.13	0.00
Improvement	1.95	2.06	0.01
Dropout	0.39	0.32	0.00
	Between First Attempt and Best Attempt		
First Grade	6.29	6.75	0.00
Max Grade	9.42	9.51	0.03
Improvement	3.12	2.75	0.00
Dropout	0.24	0.30	0.00

Note: All these numbers refer to Quiz 1. "First Grade:" grade the student received in their first attempt; "Max Grade:" maximum grade the student received; "Improvement:" improvement between first and best attempt; "Dropout:" proportion of students that did not take Quiz 2.

Table 1.2: FBS Watching the lectures

	Watchers	Non-Watchers	p value
		Before First Attempt	
First Grade	6.82	5.93	0.00
Max Grade	8.66	8.15	0.00
	Between First Attempt and Best Attempt		
Improvement	3.37	2.81	0.00
Drop-out	0.25	0.38	0.00

Note: All these numbers refer to Quiz 1. "First Grade:" grade the student received in their first attempt; "Max Grade:" maximum grade the student received; "Improvement:" improvement between first and best attempt; "Dropout:" proportion of students that did not take Quiz 2.

Table 1.3: Normal vs Distinction

	normal	distinction	p value
Grade	15.57	17.27	0.00
Procrastination	239.06	152.51	0.00
n	961.00	3971.00	

Note: All these numbers refer to Quiz 1. "Grade:" grade the student received; "Procrastination:" hours the student waited between the quiz was published and solving it.

Table 1.4: FBS taking also MWH

	Modern World	No Modern World	p value
	Quiz 1		
First Grade	7.00	6.71	0.00
Max Grade	8.71	8.60	0.18
Certificate	0.12	0.11	0.93
	Quiz 6		
First Grade	4.67	4.79	0.42
Max Grade	9.14	9.04	0.37
Certificate	0.43	0.52	0.01

Note: All these numbers refer to Quiz 1. "First Grade:" grade the student received in their first attempt; "Max Grade:" maximum grade the student received; "Certificate:" percentage of students that received a statement of accomplishment.

Table 1.5: USA vs Rest of the World

	USA	Rest of the World	p value
	Foundations of Business Strategy		
First Grade	7.08	6.66	0.00
Max Grade	8.76	8.61	0.00
Number of Attempts	2.45	2.55	0.01
Procrastination	282.18	327.34	0.00
	The Modern World: Global History since 1760		
Grade	16.01	15.28	0.00
Procrastination	254.05	320.41	0.00

Note: All these numbers refer to Quiz 1. "First Grade:" grade the student received in their first attempt; "Max Grade:" maximum grade the student received; "Number of Attempts:" number of times the student took the quiz; "Procrastination:" hours the student waited between the quiz was published and his first attempt to solve it.

Table 1.6: Number of Students that took Quiz X and Y

	Quiz 1	Quiz 2	Quiz 3	Quiz 4	Quiz 5	Quiz 6
Quiz 1	11183					
Quiz 2	6042	9106				
Quiz 3	4451	6684	6736			
Quiz 4	3623	5451	5442	5473		
Quiz 5	3167	4751	4736	4727	4775	
Quiz 6	2677	4044	4035	4031	4026	4065

Table 1.7: FBS drop-outs vs persistent students

Quiz 1 statistics	continue after Quiz 1	drop out after Quiz 1	p value
First Grade	6.98	6.42	0.00
Max Grade	9.20	7.83	0.00
Number of Attempts	2.81	2.15	0.00
Play1	16.11	20.02	0.00
Cond. Play1	27.02	22.55	0.00
PlayM	1.52	0.43	0.00
Cond. PlayM	24.68	12.24	0.00
TotalViews1	2.76	1.55	0.00
Cond. TotalViews1	7.85	5.83	0.00
Procrastination	314.96	330.77	0.00

Note: All these numbers refer to Quiz 1. "First Grade:" grade the student received in their first attempt; "Max Grade:" maximum grade the student received; "Number of Attempts:" number of times the student took the quiz; "Play1:" number of times the student pressed the play button before attempting the quiz for the first time; "Cond. Play 1:" number of times the student pressed the play button before attempting the quiz for the first time, conditioned to students who pressed the play button at least once; "PlayM:" number of times the student pressed the play button between his first and best attempt; "Cond. PlayM:" number of times the student pressed the play button between his first and best attempt, conditional on pressing it at least once; "TotalViews1:" number of threads the student view before his first attempt; "Cond. TotalViews1:" number of threads the student view before his first attempt, conditional to viewing at least one; "Procrastination:" hours the student waited between the quiz was published and his first attempt to solve it.

Table 1.8: Number of Students that took Quiz X and Y in The Modern World: Global History since 1760

	Quiz 1	Quiz 2	Quiz 3	Quiz 4	Quiz 5	Quiz 6	Quiz 7	Quiz 8	Quiz 9	Quiz 10	Quiz 11	Quiz 12	Quiz 13	Quiz 14
Quiz 1	13479													
Quiz 2	10518	10683												
Quiz 3	9115	9147	9242											
Quiz 4	8257	8282	8291	8374										
Quiz 5	7397	7417	7418	7440	7483									
Quiz 6	6958	6967	6961	6975	6966	7046								
Quiz 7	6581	6588	6584	6597	6583	6604	6676							
Quiz 8	6333	6335	6328	6334	6310	6322	6335	6437						
Quiz 9	5990	5999	5992	5997	5976	5989	6005	6035	6094					
Quiz 10	5786	5792	5786	5783	5770	5784	5797	5818	5824	5890				
Quiz 11	5562	5569	5568	5565	5551	5562	5570	5595	5597	5609	5668			
Quiz 12	5332	5340	5337	5334	5324	5332	5340	5356	5357	5365	5379	5420		
Quiz 13	5182	5189	5190	5192	5184	5191	5199	5215	5213	5215	5230	5229	5271	
Quiz 14	6958	6967	6961	6975	6966	7046	6604	6322	5989	5784	5562	5332	5191	7046

Table 1.9: MWH Drop-outs vs Persistent students

	Continue	Drop-out	p value
Grade			
Play	46.90	2.94	0.00
Cond. Play	53.03	43.76	0.00
TotalViews	3.27	0.12	0.00
Cond. TotalViews	8.47	5.32	0.00
Procrastination (hours)	268.88	447.64	0.00

Note: All these numbers refer to Quiz 1. Grade:" grade the student received; "Play:" number of times the student pressed the play button before attempting the quiz; "Cond. Play:" number of times the student pressed the play button before attempting the quiz, conditioned to students who pressed the play button at least once; "TotalViews:" number of threads the student view before solving the quiz; "Cond. TotalViews1:" number of threads the student view before solving the quiz, conditional to viewing at least one; "Procrastination:" hours the student waited between the quiz was published and solving it.

Table 1.10: Under vs over 25

	Under 25	25+	p value
Grade	16.31	16.57	0.03
Procrastination	247.80	220.71	0.03
Dropout	0.27	0.19	0.00
n	761	3776	

Note: All these numbers refer to Quiz 1. "Grade:" grade the student received; "Procrastination:" hours the student waited between the quiz was published and solving it; "Dropout:" proportion of students that did not take Quiz 2.

Table 1.11: Native vs non-native speakers

	Native speaker	Non-native speaker	p value
Grade	16.90	16.14	0.00
Procrastination	204.80	246.38	0.00
Dropout	0.19	0.22	0.01
n	2269	2240	

Note: All these numbers refer to Quiz 1. “Grade:” grade the student received; “Procrastination:” hours the student waited between the quiz was published and solving it; “Dropout:” proportion of students that did not take Quiz 2.

Chapter 2

The effects of informational nudges on students' effort and performance: Lessons from a MOOC

2.1 Introduction

Online education provides powerful new opportunities to better understand the incentives and informational environment that students face and how these affect learning. Massive Open Online Courses (MOOCs) facilitate low-cost implementation of randomized control trials; the sample sizes are large (and growing); and while many student behaviors such as time spent on homework and class participation in bricks-and-mortar classrooms have historically been costly to measure except through self and/or teacher reports, the majority of online learning activities can be observed by the researcher. While the resulting evidence on facilitating learning may help structure distance-learning environments that play an increasingly important role in human capital formation, it may also be translated for to bricks-and-mortar classrooms.

In the classroom, teachers may provide informal information about average student performance, in addition to telling students individually about their own outcomes. However, we know little about whether providing information, much less the form in which it is provided, has any effect on student effort, and probably most teachers have given little thought to how this may motivate future effort. This paper shows that very low-cost informational interventions affect how students exert effort and, ultimately, their learning success. First, I show that modest "informational nudges," as introduced by Sunstein and Thaler (2008), can affect students' behavior. Second, I examine the role of "framing," an idea emphasized by Tversky and Kahneman (1985), on how effective these nudges are. Finally, I examine how the informational nudges affect learning outcomes.

This study uses data from Coursera, a social entrepreneurship company partnering with 108 top universities worldwide to offer MOOCs.[1] As of December 2013, Coursera had 5.8 million users. Coursera users watch video lectures to learn class material, are evaluated via online quizzes, and use forums to communicate with fellow students and the instructor. I use data from the second edition of *Foundations of Business Strategy (FBS)* by Michael J. Lenox of the University of Virginia. This Coursera course enrolled 64,415 students and ran from September 9, 2013, through October 11, 2013. It has been widely reported that most MOOC students do not complete the course nor even any assignments, so for this study I focus only on students who completed the first quiz before September 15 at 3:00 p.m. Eastern Standard Time. This reduces the number of observations to 7,924, which is still a very large number considering that the third largest college freshman class in the United States has 7,740 students.[2]

Data from Coursera allows me to observe many of the inputs (e.g., lecture watch-

[1]Martinez and Diver (2014) discuss the opportunities and challenges of the data generated by Coursera.

[2]In 2008, Arizona State University had the third largest freshman class. The largest is Miami Dade College with 8,993. See Forbes: `http://goo.gl/zl1ze8`.

ing, forum participation, and number of attempts for each quiz) and outputs (e.g., grades for each attempt at each quiz, forums reputation) associated with learning. However, this is not sufficient data to understand how achievement is produced. This is because student inputs (e.g., effort) are endogenous choices. Therefore, to understand the causal effect of effort on achievement, I need an exogenous source of variation for effort. My research design generates that exogenous variation by randomly selecting a group of students to receive an email with information about their performance relative to their peers, with the goal of inducing them to alter their effort. Thus, I study the effect of both information on effort and effort on achievement.

Drawing from the economics of education and the psychology literatures, I designed two emails intended to "nudge" students to exert more effort. Sunstein and Thaler (2008) are the first to use the term nudge to describe "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any option or significantly changing their economic incentives." These emails were sent to 65% of the 7,924 students who took Quiz 1. The remaining 35% of students form the control group. The emails contain information about the relative performance of the student on Quiz 1. The two sets of emails differ only in whether they inform the students of the percentage doing better (negative framing) or worse (positive framing) than themselves.

I provide evidence that this simple email that informs students of their relative performance on the previous quiz nudges them into exerting more effort and that this effort translates, in some situations, into higher achievement. I observe positive effects on achievement for new quizzes that are given after the intervention. This suggests that students are better at adjusting by learning new material rather than relearning from the past. For example, students assigned to the negatively framed treatment (telling them the proportion of students doing better than they) attempt a quiz during the week of the intervention more often, and were ranked, on average, 8.9 percentage points higher than students in the control group. Finally, I show that the negatively framed nudge had persistent effects. In the final week of the class, students assigned to the negatively framed nudge attempted that week's quiz more often; and they ranked, on average, 5.8 percentage points higher than students in the control group.

I also show that framing plays an important role. The negative treatment tends to change outcomes for those who were doing relatively poorly, and the positive treatment tends to work for those who were doing relatively well. This indicates that the presentation of academic performance to students matters.

The remainder of this paper is organized as follows. In the next section, I discuss related research and the conceptual framework for this work. In section 2.3, I describe the data and demonstrate that the randomization worked. In section 2.4, I explain the intervention and show the results. In section 2.5, I conclude.

2.2 Related Research and Conceptual Framework

The conceptual framework for this analysis brings together ideas from economics, education, and psychology about how student time investments (effort) affect achievement, insights from psychology about the extent to which relative performance targets affect behaviors, and behavioral economics ideas about how framing motivates behavior. The objective of this inquiry is to understand whether very low cost informational interventions affect how students invest time in a course environment and, ultimately, their learning outcomes.

The term nudge was first used by Sunstein and Thaler (2008) to describe “any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any option or significantly changing their economic incentives.” For example, some firms offer employees the option of joining a program in which their saving rates are automatically increased whenever the employee gets a raise. This plan tripled saving rates in those firms. Barankay (2012) uses a randomized control trial with full-time furniture salespeople to show that removing rank feedback actuall increases sales perfamance by 11%. In this paper, I nudge Coursera students into changing their behavior by sending them information.

Gerber, Green, and Larimer (2008) show that social pressure can be used to nudge political participation. Prior to the August 2006 primary election in Michigan, approximately 80,000 households were sent one of four mailings encouraging them to vote. One experimental group received a mailing that merely reminded them that voting is a civic duty; the second group were told that researchers would be studying their turnout based on public records; a third treatment group received a mailing with the turnout record among those in the household; a fourth mailing revealed both the household’s voter turnout and their neighbors’ turnout. The effect of showing households their own voting records increases turnout by 4.9 percentage points over the control group. Showing households both their own and the voting records of their neighbors increases turnout by 8.1 percentage points compared to the control group.

It is well documented that competitive environments have causal effects on performance, both in the workplace and classroom. Mas and Moretti (2009) study whether, how, and why the productivity of a worker depends on the productivity of coworkers in the same team. They find strong evidence of productivity spillovers. Substituting a worker with below average permanent productivity with a worker with above average permanent productivity is associated with a 1 percent increase in the effort of other workers on the same shift. They find that watching others' performance and being watched can alter one's own performance. Smith (2013) uses data from the National Spelling Bee to show that when the immediate predecessor is correct, a speller has a 13 to 64 percent greater probability of making a mistake, relative to the predecessor being incorrect.

Papay, Murnane, and Willett (2011) examine how students respond to the label that they earn on the Massachusetts standardized mathematics examination (i.e., Failing, Needs Improvement, Proficient, or Advanced). Using a regression-discontinuity design, they examine the impact of the labeling by comparing the college-planning and college-enrollment decisions of students who were assigned exogenously to different labels because they scored close to, but fell on different sides of, the state-mandated labeling cut-points. They find that earning a more positive performance label causes urban, low-income students to attend college at greater rates. They provide two explanations for this effect. First, cognitive limitations may make interpretations of complicated test-score data difficult and may increase students' reliance on the performance labels. Second, the labels may evoke emotional responses.

Fryer (2013) conducted a randomized field experiment in Oklahoma City public schools, which provided information to students on the link between human capital and future outcomes such as unemployment, incarceration, and wages. The essential element of the experiment was a cellphone that was provided to 1,470 students in the treatment group. Students received one text message per day containing this information. Three facts emerge from this study: (1) students update their beliefs about the returns to education in response to the text messages (2) students report that they are putting more effort into their work, and (3) there are no detectable changes in academic achievement. Fryer argues that the explanation for this is that students do not fully understand the education production function. Earlier work by Fryer shows that paying young people to finish reading books (that is, inducing them to invest in inputs) has a bigger effect than incentives to do well on exams; see Fryer (2011). Therefore, to the extent that informational interventions affect productivity-enhancing behavior, such interventions might improve learning outcomes.

Numerous studies find that students' effort is positively correlated with academic achievement; see for example Carbonaro (2005); Johnson, Crosnoe, and Elder Jr (2001). However, it is very difficult to disentangle the roles of pedagogical methods, students' effort, and individual characteristics (including prior preparation and innate ability) in the measurement of achievement. One reason is that these studies rely on student-reported or teacher-reported effort. Yet, self-assessment may be subject to substantial misreporting, which is, in turn, correlated with underlying student characteristics, resulting in biased estimates. Another problem is that effort is an endogenous choice. I designed a scalable very low cost intervention that generates a source of exogenous variation in an attempt to enable me to estimate causal effects of effort.

Tversky and Kahneman (1985) present evidence showing how seemingly inconsequential changes in the formulation of choice problems caused significant shifts of preferences. Zhang and Buda (1999) study positive and negative framing by the advertising industry. They find that framing has a significant influence on consumer responses to advertisements. They also find that this effect is more salient for people with a low need for cognition than those with a high need for cognition.[3] Framing effects can be viewed as heuristic errors. That is, if people are boundedly rational, then the presentation of a choice may draw attention to new aspects of a problem, leading to people to make mistakes in pursuing their true underlying preferences (Rabin, 1998). If this is the case, students may react differently to the information that they are performing better than 20% of their classmates than to the information that they are doing worse than 80% of their classmates, even though this could be interpreted as violating the assumption of rationality. Levitt et al. (2012) conducted a series of field experiments involving thousands of primary and secondary school students to explore this. They find that incentives framed as losses have more robust effects than comparable incentives framed as gains.

Although MOOCs are relatively new, research interest in them is rapidly growing. In a systematic study of published literature, Liyanagunawardena, Adams, and Williams (2013) found that through 2012, there had only been 45 distinct articles where MOOCs or their use are the primary focus. However, the pattern of publication shows a quickly increasing trend (from one in 2008 to 26 in 2012). Of these, only 15 were classified under the broad heading of "Educational Theory." In 2013, research using MOOCs has continued, but it is evidently still in an early state. The nascent research is still trying to understand who is enrolling in MOOCs and why (see, for example, Christensen et al. (2013)), and overall trends in completion and engagement (see, for example, Kizilcec, Piech, and Schneider (2013)). Those researchers that have made use of the MOOCs' data have typically limited their focus to classification or

[3] The need for cognition in psychology is a personality variable reflecting the extent to which individuals are inclined towards effortful cognitive activities. The need for cognition has been variously defined as "a need to structure relevant situations in meaningful, integrated ways" and "a need to understand and make reasonable the experiential world"; see Cohen, Stotland, and Wolfe (1955) and Cacioppo and Petty (1982).

broad associative observations. Kizilcec, Piech, and Schneider (2013), for example, present a simple classification method that identifies a small number of longitudinal engagement trajectories in MOOCs. A limitation of descriptive studies is that they can detect association among variables, but they cannot rule out the possibility that the association was caused by an omitted factor that is correlated with outcomes. For example, it is possible that some unobserved characteristic, such us being a hard worker, explains both forum participation and course completion, rather than forum participation increasing course completion. To overcome this selection bias, I use a randomized controlled trial. This is a rigorous way of determining whether a causal relation exists between effort and achievement outcomes.

2.3 Data

2.3.1 Course Design

Data from Massive Open Online Courses (MOOCs) provide an unprecedented opportunity to get inside the black box of student learning. These data include time-stamped logs of student activities such as viewing lectures, submission of assignments; participating in forums; clickstream logs (which track user activity on the course website); page views and lecture video interaction (e.g., video seek events); geolocation information from Internet protocol addresses (IP addresses); all the courses a student is currently and has previously taken; and student background surveys. Additionally, survey data provide some demographic characteristics such as age, and level of education.

I use data from the second edition of *Foundations of Business Strategy* (FBS) by Michael J. Lenox of the University of Virginia. This Coursera course had 64,415 students enrolled and ran from September 2, 2013 to October 11, 2013. FBS explores the underlying theory and frameworks that provide the foundations of a successful business strategy. The class is divided into weekly modules. Each weekly module consists of an introductory video, a reading from the strategist's toolkit, a series of video lectures, a quiz, and a case study to illustrate points in the lectures. Students wishing to receive a statement of accomplishment must satisfy the following criteria:

1. Complete six quizzes. Students can take each quiz as many times as they want. The score of record is the best score for each quiz. Quizzes have 10 questions with four choices each.

2. Submit a final project (a strategic analysis of an organization of their choosing).

3. Assess five peers' strategic analysis using the peer assignment function.

Final grades are out of 100 points. Of those, 50 points are for the final project, 42 points for the quizzes (spread evenly over the six quizzes), and 8 points for post and comment up-votes. Those who received 70 points or more and assessed five peers' strategic analysis received a statement of accomplishment.

In this paper, I use the 7,924 students who completed Quiz 1 before September 15 at 3:00 p.m. EST. Of these students only 1,539 ended up receiving a statement of accomplishment. Figure 2.1 shows when each of the quizzes was published.

All quizzes share the same deadline of October 15; students can take each quiz as many times as they want; and the score of record is the best score on each quiz. Therefore, although the intervention happened at the beginning of the third week, students can retake Quizzes 1 or 2 if they wish.

2.3.2 Demographic Characteristics

I am able to determine students' location from their IP address. As shown in Figure 2.2, most students are from the United States (1,997), followed by India (879), Brazil (413), and the United Kingdom (296).[4]

As shown in Martinez and Diver (2014), most Coursera students do not complete surveys, so their demographic characteristics are not well known. To get some insight about other demographic characteristics, I combined data from the course survey with Coursera's standardized demographic survey.

Out of the 7,924 students, only 3,025 answered the question about their education.[5] Figure 2.3 shows the education distribution for the respondents.

I am able to observe age for only 2,330 students. As shown in Figure 2.4, the youngest student in the course was 14, the median 30, and the oldest 76.

Finally, gender data is available for only 1,079 students. Figure 2.5 shows that most students, who answered the survey, are male.

Although this seems to indicate that these students were well-educated young men in their 30s, it is possible that less well-educated people are less likely to complete the survey. As I discuss in Martinez and Diver (2014), current survey data from MOOCs should be used with caution.

[4]For the US, IP geolocation allows me to know the metropolitan area for each student. The state with the largerst number of students is California (302), followed by Virginia (186), and New York (161). Interactive maps are available online at http://goo.gl/NGJM4g.

[5]Out of the 64,415 students enrolled on this course, only 4,548 answered the course survey, and 6,699 the standardized survey.

2.3.3 Treatment and Control Balance

In this section, I show evidence that the randomization worked as intended. In the next section, I will describe the randomized treatments, in which 35% of the sample were in the control group and half of the treatment group received the positive treatment and the other half the negative treatment. In a randomized controlled trial such as this, the econometrics entail fairly simple comparisons between the treatment and control group, so long as the groups were randomly selected.

Given the use of a random number generator to assign students to the treatment or control status, from the large number of students and the law of large numbers, the observable and unobservable characteristics should be the same across groups. Table 2.1 summarizes a few learning-related variables for the treatment and control groups.[6] "Best Grade Before Experiment" is the grade out of 10 points for the best attempt a student achieved for Quiz 1 before the experiment. "Clicks on Play Before 1st Attempt" is the number of times the student presses the play button on the video player to watch a lecture. "Forum Thread Views Before 1st Attempt" is the number of thread views a student made on the course forums before attempting the quiz for the first time.

Table 2.1 shows coefficient estimates from regressions of each of these variables on the treatment dummies. As expected, the coefficients are very close to zero and are not statistically significant.

[6] Recall that I only have basic demographic information for 14% to 38% of the sample depending on the question. Nevertheless, these variables are also balanced between the treatment and control groups.

2.4 Experimental and Results

2.4.1 Experimental Design

I designed the following experiment to disentangle the role of information and framing on students' choices, and ultimately the role of these choices on achievement. The experiment consists of nudging students with an email containing information about their relative class performance. The nudge is framed in two different ways to allow me to test the hypothesis that students react differently to positive and negative framing.

A day after the third FBS quiz was published, on September 16, 2013, I sent emails to 65% of the 7,924 students who had taken Quiz 1. The remaining 35% of students form the control group. The two sets of emails differ only in whether they informed the students of the percentage doing better (negative framing) or worse (positive framing) than themselves. These emails were sign as "University of Virginia MOOC Research Team" and read as follows:

> Subject: [Foundations of Business Strategy] Quiz 1
>
> *Dear [name],*
>
> *This information about your performance may benefit you. You obtained a [maxGrade] on the first quiz. That means that you are doing [better] / [worse] than [%worse] / [%better] of the class.*
>
> *Best,*
>
> *University of Virginia MOOC Research Team*
>
> *PS: This email was generated with data from Sunday September 15 at 3:00pm EST.*

I sent these emails via sendgrid.com, which provides additional analytics on whether the emails are read. The confirmed unique open rate for the positively framed email was 38.33% with six bounces and one spam report. For the negatively framed email, the confirmed unique open rate was 39.39% with eight bounces and zero spam reports, so the open rates are not statistically different from each other. To get the unique open rate, sendgrid inserts a white pixel in the body of the email. Because images are blocked by default on many email clients, these numbers are lower bounds for the open rates. The open rate is crucial to calculating the effect of the treatment on the treated.

Next, I discuss how the treatments affected the following outcomes for each student that reflect both effort and achievement:

- Go Back: the decision to retake Quiz 1 after September 16
- Attempts: the number of attempts the student makes for a given quiz
- Best Grade: the maximum grade the student received
- Ranking: the percentile ranking for the student[7]

People may respond by changing their effort on any or all of the quizzes. I expect the effects of the experiment to be different across quizzes for two reasons. First, the intervention only gives information about relative performance for Quiz 1. Since the informational nudge makes Quiz 1 salient, the effects on Quiz 1 outcomes and choices

[7] "Go Back" is only relevant for Quizzes 1 and 2 because these were published before the intervention. For the rest of the quizzes, I use "Attempts" as a measure of effort.

Although it would be possible to compare the effects of the nudge on variables such as number of clicks on the play button or number of threads read, there is no obvious interpretation for these variables. Clicking fewer times on the play button or reading fewer threads does not imply less time doing those activities. Alas, Coursera does not collect data on time use at the moment. Martinez and Diver (2014) discuss this in more detail.

should be greater than the effects on other quizzes. Second, the emails were sent the day after Quiz 3 was published. Thus, the effects on Quiz 3 may also be greater.

2.4.2 Experimental Effects

In this section I examine the effects of the intervention on students choices (i.e., go back and retake a quiz after the intervention, how many times to attempt a quiz, and whether or not to drop out) and outcomes (i.e., student ranking on each quiz and course completion). I find that a very-low cost intervention can nudge students into exerting more effort and that this effort translates, especially for later quizzes, into higher achievement. I show evidence that the framing of the nudge matters. Students who were doing relatively poorly respond to the negatively-framed nudge with more effort, with fairly big improvements in test scores especially for those who were not already falling behind on taking quizzes. On the other hand, students who were doing relatively well respond to the positively-framed nudge.

Effects on Quiz 1

The informational nudge was sent before Quiz 3 but makes Quiz 1 outcomes salient. For example, I find that when restricting the sample to students who did not already have a perfect score before the intervention, the negatively-framed nudge increases the probability of retaking Quiz 1 by 40%. In column 4 of Table 2.2 the negative treatment has a statistically significant effect on "Go Back", the proportion of students who went back to take Quiz 1. 8.7% of these students went back to Quiz 1 after the email was sent while only 7.1% of students in the control group did. The positive treatment, however, does not have a statistically significant effect.

The distribution of grades before the intervention, in Figure 2.6, shows a lot of students bunched at 10. These students already have a perfect score before the treatment and therefore no room for improvement. Next, I exclude the perfect scores from the sample in Table 2.3. Compared to an estimated effect of 1.6 percentage points on retaking Quiz 1 for the entire sample, the negatively-framed had a statistically significant estimated effect of 4.0 percentage points, or 40%, which is quite large. Meanwhile, the positive treatment had an effect of 1.7 percentage points, which falls a little short of statistical significance at the 90% level. These results indicate that informational nudges affect effort, and that negative information has a more powerful effect. The negatively-framed treatment has an estimated effect on retaking Quiz 1 that is 2.3 percentage points greater than the positively framed treatment.

By choosing to go back and retake Quiz 1, students are exerting more effort, and "Best Grade" on Quiz 1 rose, with an insignificant estimate for the positive treatment and an estimate of 0.10 that falls just short of statistical significance (p-value of 0.1044) for the negative treatment. This may support the hypothesis that students do not fully understand the education production function, as some exerted additional effort without gaining a better outcome, and motivates future research in which the nudge guides students about how to exert effort (i.e. spend more time in forums, or watching videos).

Effects on Quiz 2

Of note, Quiz 2 was released before the treatment, though scores on Quiz 2 were not mentioned in the email. Following the same logic as before, a student could decide to go back to Quiz 2 in order to improve his ranking. This might be less costly than going back to Quiz 1 because less time had passed since the material had been covered. However, students have less information about their ranking for Quiz 2.

In Table 2.4, I summarize the effects of treatment participation on Quiz 2. In order to avoid concerns about attrition from the course after the delivery of the treatment, students who took Quiz 1 but not Quiz 2 were included with a grade of zero. The positively-framed treatment had a statistically significant estimated effect of 2.9 percentage points on retaking Quiz 2. Interestingly, this positive treatment was more effective on students who had a perfect score on Quiz 1 (column 5). Students with a perfect score on Quiz 1 and who were assigned to the positively-framed nudge are 4.1 percentage points more likely to retake Quiz 2. This perhaps suggests that the nudge revealed that a perfect score is not something extraordinary but the norm.[8]

Effects on Quiz 3

Performance on Quiz 3, which was published the day before the intervention, provides a strong indication of whether it affected the forward-looking behavior of participants. If a student wants to improve his performance, it might be ideal to exert more effort on Quiz 3.

Including students who did not take Quiz 3 dilutes the effects of the intervention. Columns 7 and 8 from Table 2.4 show that there is no statistically significant effect on whether or not a student took Quiz 3. Because of this, I will restrict my analysis to those who took Quiz 3.

Figure 2.7 shows that students who took Quiz 3 and were assigned to one of the treatment groups scored significantly higher than students in the control group. Column 1 of Table 2.5 shows that, on average, students assigned to any of the two

[8] 45% of the students had a perfect score on Quiz 1 before the intervention, after the intervention 48%.

treatments were ranked 2.7 percentage points better on Quiz 3 than students in the control group. Once again, as shown in column 2, the negatively-framed treatment was more effective, as students in assigned to this group were ranked 3.4 percentage points better. Column 3 shows that the negative treatment is more effective for students who performed poorly on the first quiz before the intervention. For example, a student who scored less than 8 and was assigned to the negative treatment was ranked, on average, 7.9 percentage points better than a student in the negative treatment but with a perfect score, while a student who scored 8 or 9 was ranked only 5.2 percentage points better than a student with a perfect score in Quiz 1. Columns 4 and 5 show that these improvements in ranking are at least partially explained by an increase in effort. On average, a student assigned to the negatively-framed treatment attempted Quiz 3 0.154 more times than a student in the control group, while a student in the positively-framed group attempted the quiz 0.119 more times than a student in the control group. In columns 6 and 7, I show that the treatments have no statistically significant effect on whether the students took Quiz 3 before Quiz 4 was published (are "on track").

Martinez and Diver (2014) show that students who procrastinate are different from students who do not. Martinez (2014b) goes beyond establishing this correlation and shows a causal relationship that procrastination causes low achievement. To examine further whether the informative nudges in this paper have an heterogeneous effect for these different types of students, in the next table I further restrict the sample to students on track in order to focus on those who do not procrastinate (they took Quiz 3 on time), since procrastination is not affected by the treatment.

In Table 2.6, I show that the nudges are even more effective for students who are on track, particularly for those who did not have a perfect score in the first Quiz before the intervention. For example, students assigned to the negatively framed treatment who did not have a perfect score were ranked, on average, 8.9 percentage points better than students in the control group. These students attempted the Quiz, on average, 0.5 times more than students in the control group.

Effects on Quizzes 4 to 6

Quiz 4 was published a week after the intervention. Changes in choices and outcomes for this quiz provide information about the persistence of the nudge. Persistent effects of the nudge may show that the information itself remains salient or that being nudged to learn earlier material helps students with later material.[9]

Table 2.7 summarizes the effects for Quiz 4. Students on track who scored 9 or less on Quiz 1 attempted Quiz 4, on average, 0.397 times more than students in the control group. Their ranking was 7.2 percentage points higher than students in the control group.

Table 2.8 summarizes the results for Quiz 5. The positively framed treatment increased the probability of being on track by 7.6 percentage points. Students on track who did not have a perfect score on Quiz 1 before the intervention were ranked 7.3 percentage points higher than students in the control group on Quiz 5.

Three weeks after the nudge, once Quiz 6 is published, the effects of the negatively framed nudge persist. Table 2.9 shows that students assigned to the negatively-framed nudge work harder and perform better on Quiz 6. These students attempt the quiz, on average, 0.235 more times, and are ranked 5.8 percentage points higher than students in the control group. Finally, the intervention had no statistically significant effect on

[9] In a conversation about these results, Professor Lenox told me: “the course builds on itself, but I think it is fair to say that someone can do well on a quiz by simply watching that week’s lectures.”

the final project, which accounts for 50% of the final grade. Probably because of this, the intervention had no effect on whether a student obtained the course certificate.

2.5 Conclusions

This paper main contribution is to show that informational nudges can change students' behavior and improve their achievement. The nudges raise achievement on new quizzes given after the intervention. Some students also retake old quizzes, but without improvements in scores, suggesting that it is better for students to make adjustments when learning new material rather than relearning older material. For instance, I find that, among students who were on track (taking a quiz before the next one was published) but who had not obtained a perfect score on Quiz 1, those who were assigned to receive the negatively framed email were ranked, on average, 8.9 percentage points higher on Quiz 3 than students in the control group.[10] This represents a major impact per dollar spent. This also shows that students care about relative performance in a competitive environment, otherwise they would not respond to a simple informational email. While these experiments are conducted using a MOOC, low-cost informational interventions could similarly improve achievement in traditional classrooms.

The fact that an increase in effort does not always translate into higher achievement, as I find here especially for increases in effort on previously published quizzes, supports the hypothesis that students do not fully understand the education production function. In future research, I will use more directed interventions to test this

[10] Almost 20% of the students who took Quiz 3 were in this sub-population for negatively framed treatment.

hypothesis further. For example, the emails will recommend specific behavior (e.g., taking notes, reading the forums, etc) or will point out the rewards for acting early and avoiding procrastination.

As important, I show what matters when nudging students. The negative treatment tends to change outcomes for those who were doing relatively poorly, and the positive treatment tends to work for those who were doing relatively well. This suggests that the presentation of academic performance matters.

Finally, some questions remain unanswered. For example, why do students care about their grade? Once they are above the minimal threshold they all get the same certificate. The answer to this question is probably a mixture of caring about learning and competitive spirit. It is also possible that the email generates a psychological effect that makes students concerned about monitoring or reputation. This is something that I plan to test in future research.

Figure 2.1: Time Line

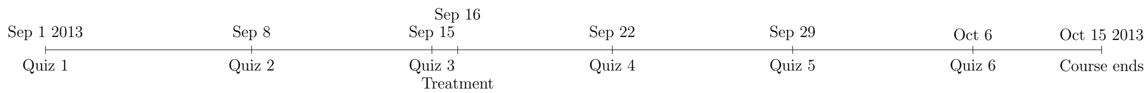


Figure 2.2: Students' geolocation

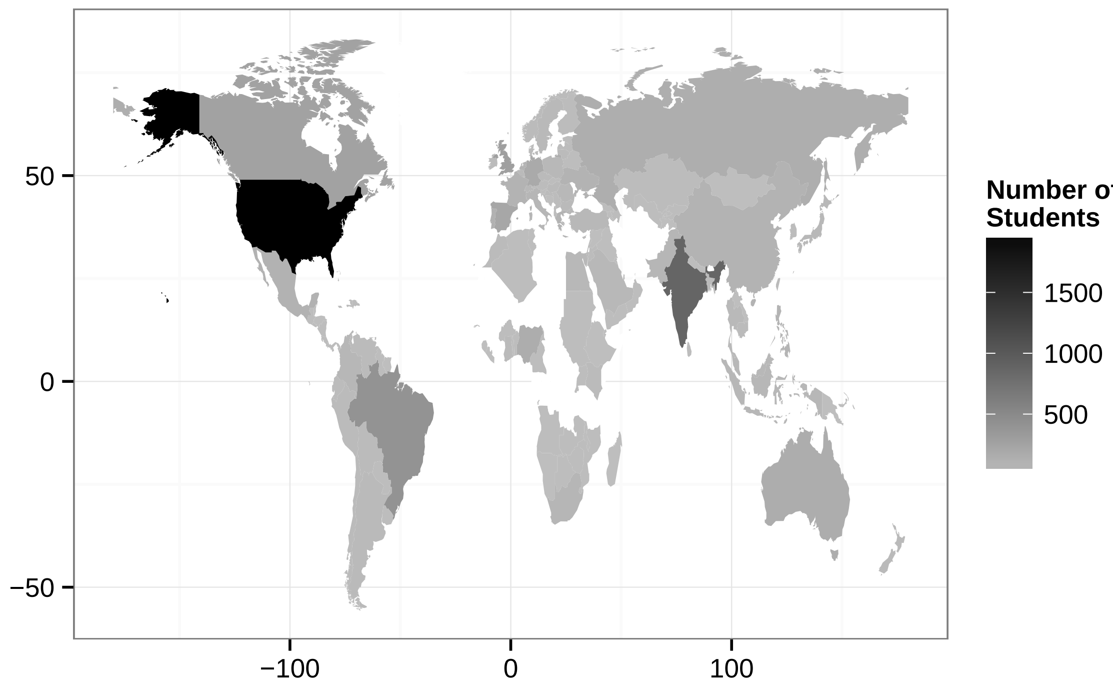


Figure 2.3: Please indicate your highest level of education (survey question)

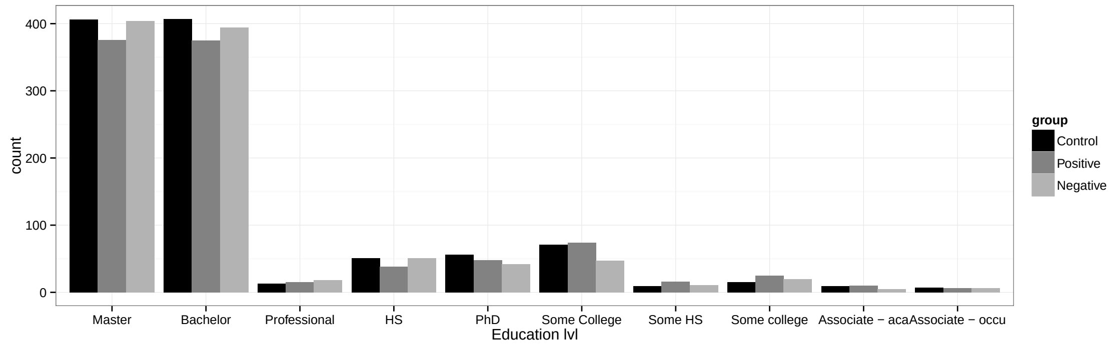


Figure 2.4: Please indicate your age (survey question)

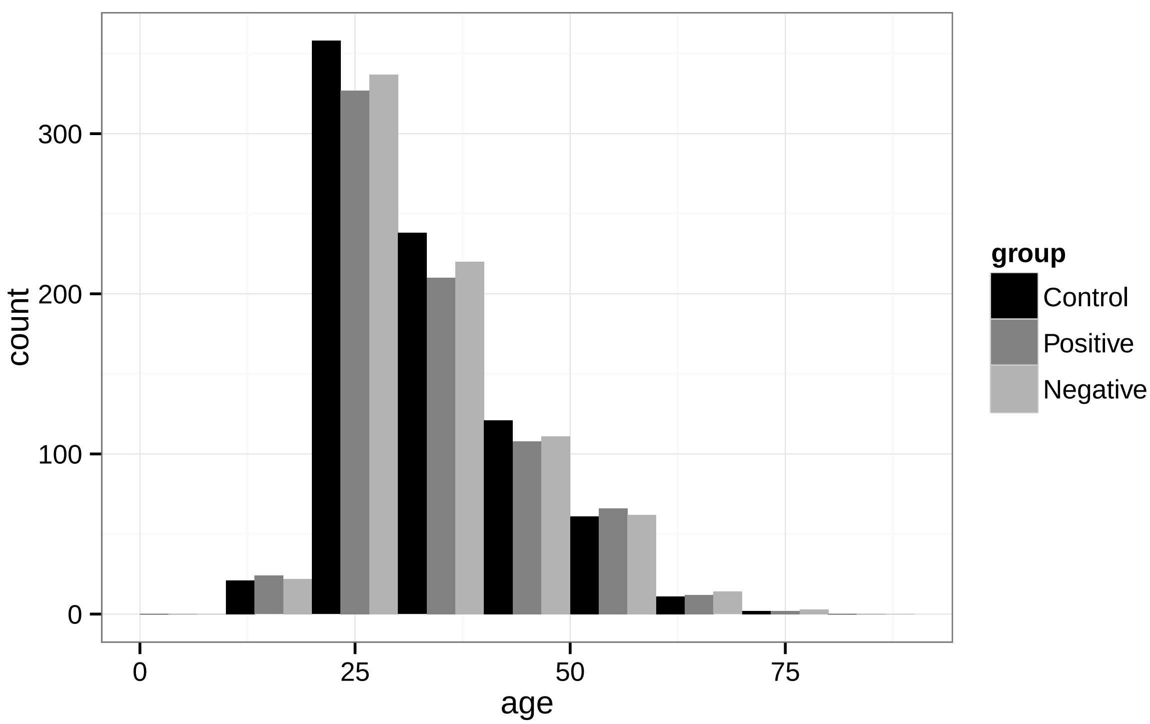


Figure 2.5: What is your gender? (survey question)

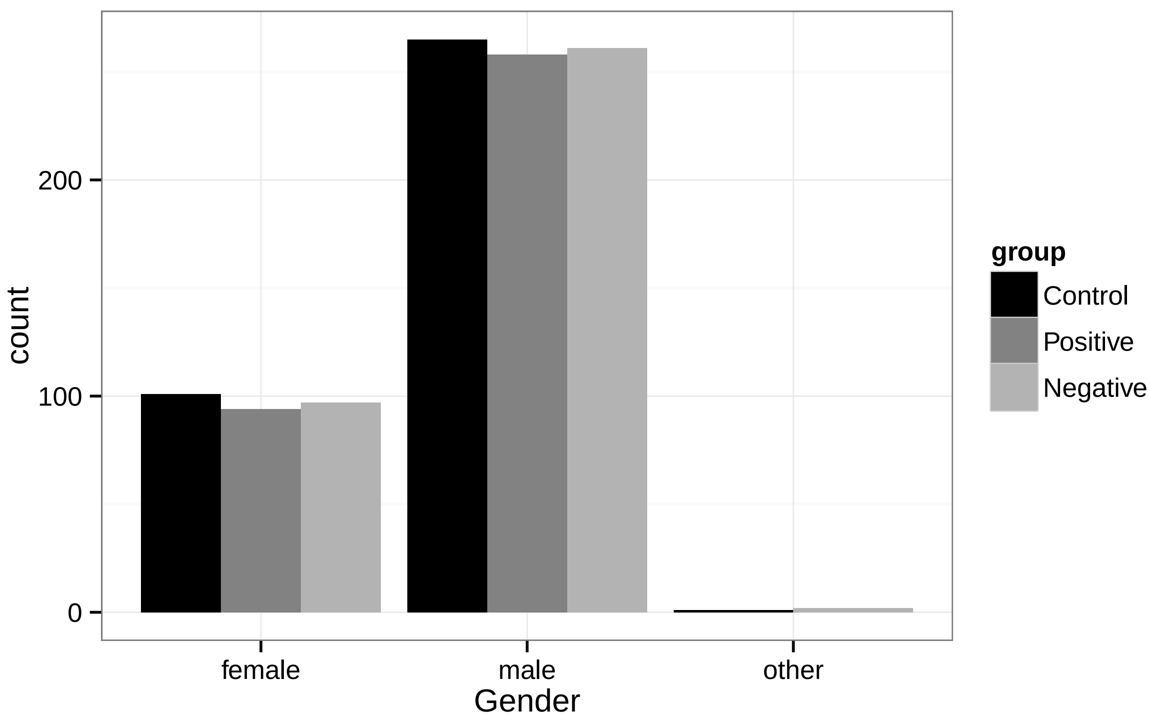


Figure 2.6: Grade distribution before the intervention of Quiz 1 by group

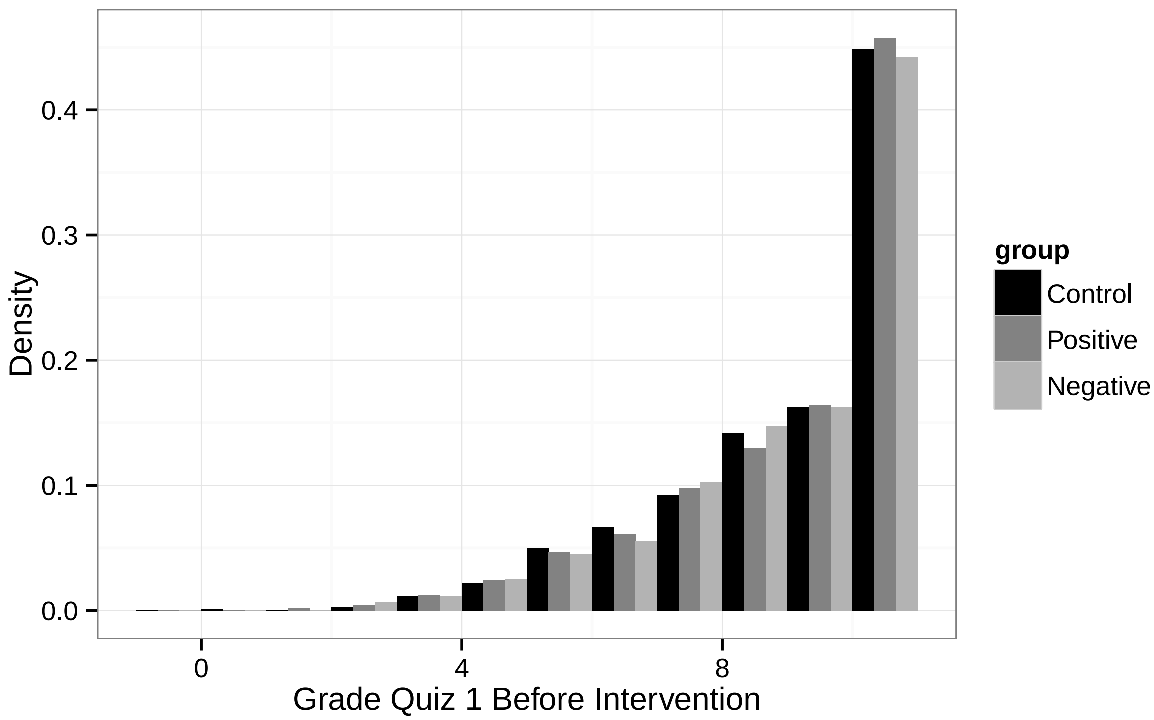
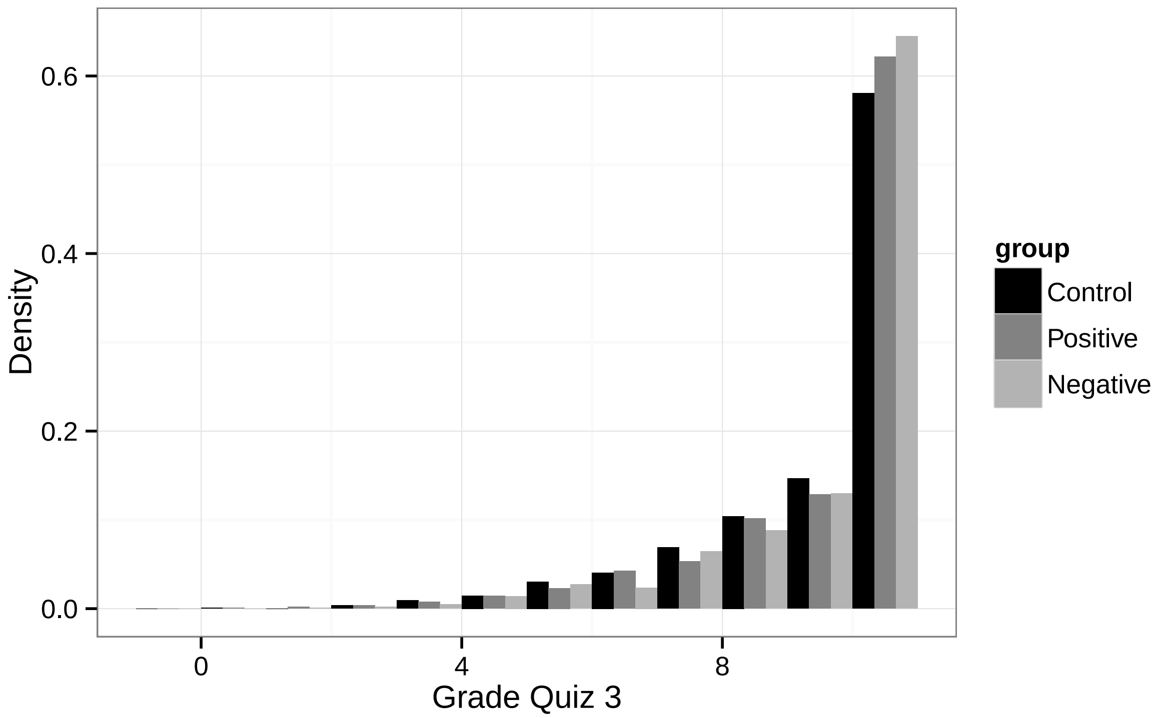


Figure 2.7: Quiz 3 grades distribution



Note: An interactive version of this graph is available at `http://goo.gl/nA4mKM`

Table 2.1: Treatment and Control Balance

	Dependent variable:					
	Best Grade Before Experiment		Clicks on Play Before 1st Attempt		Forums Threads Views Before 1st Attempt	
	(1)	(2)	(3)	(4)	(5)	(6)
Any Treatment	0.001		−0.547		−0.049	
	(0.043)		(0.685)		(0.245)	
Positive		0.014		−0.297		−0.127
		(0.050)		(0.795)		(0.284)
Negative		−0.012		−0.795		0.029
		(0.050)		(0.794)		(0.284)
Constant	8.509***	8.509***	20.180***	20.180***	2.238***	2.238***
	(0.035)	(0.035)	(0.555)	(0.555)	(0.198)	(0.198)
Observations	7,924	7,924	7,924	7,924	7,924	7,924

Note: *p<0.1; **p<0.05; ***p<0.01
Students were assigned to the treatment or control group using the default random number generator in R.

Table 2.2: Effect of Treatment Participation on Quiz 1, Full Sample

	Dependent Variable:							
	Best Grade		Go Back		Took Quiz 2		Dropouts' Best Grade	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Any Treatment	0.046		0.010		0.012		0.024	
	(0.042)		(0.006)		(0.012)		(0.071)	
Positive		0.050		0.003		0.015		0.066
		(0.049)		(0.007)		(0.014)		(0.083)
Negative		0.041		0.016**		0.009		−0.016
		(0.049)		(0.007)		(0.014)		(0.083)
Constant	8.602***	8.602***	0.071***	0.071***	0.537***	0.537***	7.973***	7.973***
	(0.034)	(0.034)	(0.005)	(0.005)	(0.010)	(0.010)	(0.058)	(0.058)
Observations	7,924	7,924	7,924	7,924	7,924	7,924	3,606	3,606

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2.3: Effect of Treatment on Quiz 1, Restricted to Grade< 10 Before Nudge

	Dependent Variable:							
	Best Grade		Go Back		Took Quiz 2		Dropouts' Best Grade	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Any Treatment	0.077		0.029***		0.008		0.064	
	(0.055)		(0.010)		(0.016)		(0.077)	
Positive		0.050		0.017		0.008		0.101
		(0.064)		(0.012)		(0.018)		(0.090)
Negative		0.103		0.040***		0.007		0.027
		(0.063)		(0.012)		(0.018)		(0.089)
Constant	7.464***	7.464***	0.100***	0.100***	0.439***	0.439***	6.967***	6.967***
	(0.044)	(0.044)	(0.008)	(0.008)	(0.013)	(0.013)	(0.062)	(0.062)
Observations	4,362	4,362	4,362	4,362	4,362	4,362	2,424	2,424

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2.4: Effect of Treatment Participation on Quizzes 2 and 3

	Dependent Variable:									
	Best Grade		Go Back				Took Quiz 3		Dropouts' Best Grade	
	All	All	All	All	Perfect Quiz 1	Not Perfect Quiz 1	All	All	Dropouts	Dropouts
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Any Treatment	0.132		0.020*				0.005		0.113	
	(0.109)		(0.011)				(0.012)		(0.108)	
Positive		0.132		0.029**	0.041**	0.019		0.002		0.171
		(0.126)		(0.013)	(0.019)	(0.017)		(0.013)		(0.125)
Negative		0.133		0.012	0.004	0.018		0.008		0.055
		(0.126)		(0.013)	(0.019)	(0.017)		(0.013)		(0.125)
Constant	4.698***	4.698***	0.302***	0.302***	0.327***	0.282***	0.397***	0.397***	1.801***	1.801***
	(0.088)	(0.088)	(0.009)	(0.009)	(0.014)	(0.012)	(0.009)	(0.009)	(0.087)	(0.087)
Observations	7,924	7,924	7,924	7,924	3,562	4,362	7,924	7,924	4,749	4,749
Note:										*p<0.1; **p<0.05; ***p<0.01

Table 2.5: Effect of Treatment Participation on Quiz 3

	Dependent Variable:						
	Ranking			Attempts		On Track	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Any Treatment	0.027***			0.137**		0.010	
	(0.009)			(0.056)		(0.019)	
Positive		0.020*	0.010		0.119*		0.009
		(0.011)	(0.012)		(0.065)		(0.022)
Negative		0.034***	0.006		0.154**		0.012
		(0.011)	(0.012)		(0.065)		(0.022)
$Q1 < 8$			-0.353^{***}				
			(0.020)				
$10 > Q1 \geq 8$			-0.233^{***}				
			(0.015)				
Positive x $Q1 < 8$			0.018				
			(0.028)				
Negative x $Q1 < 8$			0.079***				
			(0.028)				
Positive x $10 > Q1 \geq 8$			0.012				
			(0.022)				
Negative x $10 > Q1 \geq 8$			0.052**				
			(0.021)				
Constant	0.482***	0.482***	0.600***	2.752***	2.752***	0.474***	0.474***
	(0.008)	(0.008)	(0.009)	(0.046)	(0.046)	(0.015)	(0.015)
Observations	3,175	3,175	3,175	3,175	3,175	3,175	3,175
Note:							*p<0.1; **p<0.05; ***p<0.01

Table 2.6: Effect of Treatment Participation on Quiz 3, Restricted Sample

	Dependent variable:							
	Ranking				Attempts			
	On Track	On Track	On Track and Not Perfect Quiz 1	On Track and Not Perfect Quiz 1	On Track	On Track	On Track and Not Perfect Quiz 1	On Track and Not Perfect Quiz 1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Any Treatment	0.023*		0.057**		0.213***		0.394***	
	(0.013)		(0.023)		(0.076)		(0.125)	
Positive		0.021		0.022		0.223**		0.269*
		(0.015)		(0.026)		(0.088)		(0.146)
Negative		0.026*		0.089***		0.204**		0.512***
		(0.015)		(0.026)		(0.087)		(0.143)
Constant	0.511***	0.511***	0.316***	0.316***	2.691***	2.691***	2.059***	2.059***
	(0.011)	(0.011)	(0.018)	(0.018)	(0.062)	(0.062)	(0.102)	(0.102)
Observations	1,526	1,526	552	552	1,526	1,526	552	552
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 2.7: Effect of Treatment Participation on Quiz 4

	Dependent variable:							
	Took Quiz 4		Took Quiz 4 Before 5		Ranking		Attempts	
	All	All	Took Quiz 4, Not Perfect Quiz 1	Took Quiz 4, Not Perfect Quiz 1	Took Quiz 4, Not Perfect Quiz 1, and On Track	Took Quiz 4, Not Perfect Quiz 1, and On Track	Took Quiz 4, Not Perfect Quiz 1, and On Track	Took Quiz 4, Not Perfect Quiz 1, and On Track
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Any Treatment	0.006		0.005		0.036		0.302^{**}	
	(0.011)		(0.021)		(0.028)		(0.126)	
Positive		0.008		0.012		0.0002		0.208
		(0.013)		(0.024)		(0.032)		(0.145)
Negative		0.005		-0.002		0.072^{**}		0.397^{***}
		(0.013)		(0.024)		(0.032)		(0.146)
Constant	0.326^{***}	0.326^{***}	0.488^{***}	0.488^{***}	0.477^{***}	0.477^{***}	2.040^{***}	2.040^{***}
	(0.009)	(0.009)	(0.017)	(0.017)	(0.023)	(0.023)	(0.104)	(0.103)
Observations	7,924	7,924	2,613	2,613	453	453	453	453
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 2.8: Effect of Treatment Participation on Quiz 5

	Dependent Variable:							
	Took Quiz 5		Took Quiz 5 Before 6		Ranking		Attempts	
	All	All	Took Quiz 5, Not Perfect Quiz 1	Took Quiz 5, Not Perfect Quiz 1	Took Quiz 5, Not Perfect Quiz 1, and On Track	Took Quiz 5, Not Perfect Quiz 1, and On Track	Took Quiz 5, Not Perfect Quiz 1, and On Track	Took Quiz 5, Not Perfect Quiz 1, and On Track
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Any Treatment	0.004		0.055^{**}		0.023		-0.132	
	(0.011)		(0.022)		(0.028)		(0.153)	
Positive		0.004		0.076^{***}		-0.023		-0.210
		(0.012)		(0.025)		(0.031)		(0.173)
Negative		0.004		0.035		0.073^{**}		-0.049
		(0.012)		(0.025)		(0.032)		(0.176)
Constant	0.286^{***}	0.286^{***}	0.510^{***}	0.510^{***}	0.484^{***}	0.484^{***}	2.277^{***}	2.277^{***}
	(0.009)	(0.009)	(0.018)	(0.018)	(0.023)	(0.023)	(0.128)	(0.128)
Observations	7,924	7,924	2,285	2,285	454	454	454	454
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 2.9: Effect of Treatment Participation on Quiz 6

	Dependent variable:									
	Took Quiz 6		Ranking		Attempts		Final Project		Certificate	
	All	All	Took Quiz 6, and Not Perfect Quiz 1	Took Quiz 6, and Not Perfect Quiz 1	Took Quiz 6, and Not Perfect Quiz 1	Took Quiz 6, and Not Perfect Quiz 1	Took Quiz 6, and Not Perfect Quiz 1	Took Quiz 6, and Not Perfect Quiz 1	Took Quiz 6, and Not Perfect Quiz 1	Took Quiz 6, and Not Perfect Quiz 1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Any Treatment	0.003		0.021		0.087		0.027		-0.015	
	(0.010)		(0.021)		(0.109)		(1.676)		(0.037)	
Positive		0.004		-0.016		-0.063		-0.277		-0.054
		(0.012)		(0.024)		(0.126)		(1.986)		(0.043)
Negative		0.002		0.058^{**}		0.235^{*}		0.299		0.024
		(0.012)		(0.024)		(0.126)		(1.929)		(0.043)
Constant	0.253^{***}	0.253^{***}	0.486^{***}	0.486^{***}	2.704^{***}	2.704^{***}	70.380^{***}	70.380^{***}	0.458^{***}	0.458^{***}
	(0.008)	(0.008)	(0.017)	(0.017)	(0.088)	(0.088)	(1.340)	(1.342)	(0.030)	(0.030)
Observations	7,924	7,924	802	802	802	802	483	483	802	802
Note:										*p<0.1; **p<0.05; ***p<0.01

Chapter 3

Never put off 'till tomorrow?

3.1 Introduction

Can patterns of behavior and actions be manipulated to improve students' outcomes? Conventional wisdom establishes that procrastination is bad for student achievement: without deadlines students may procrastinate on their work, and this may reduce learning.

Procrastination is difficult to measure and endogenous to most outcomes of interest. Most papers rely on self-reported procrastination, which may cause a Hawthorne effect: students may change their behavior and procrastinate less because they are asked to record their behavior. Also, low ability students may be more likely to procrastinate, and changing their work habits might not improve their outcomes.

Massive Online Open Courses (MOOCs) provide an ideal laboratory to study procrastination. The rich data that MOOC providers collect include when the course material is published and when the students interact with it. Therefore, researchers can observe procrastination directly without relying on self reported measures. Using data from Coursera, a MOOC provider, Martinez and Diver (2014) provide descriptive evidence of the strong negative correlation between procrastination and achievement, as shown in Figure 3.1: students who attempt Quiz 1 for the first time later, perform on average worse than those who do not procrastinate.

In this paper, I use two approaches to estimate the causal effect of procrastination on achievement. First, because MOOCs collect information on individual IP addresses, I can use weather data for an instrumental variables (IV) approach. Second, I use directive nudges for an experimental approach.

Weather shocks provide a source of variation that predicts procrastination. I show that rain and snow affects when a student takes a quiz, and therefore can be use as an IV. For example, a student is 2.3 percentage points less likely to attempt the first quiz the day it is published on a day with with rainfall, but 5.0 percentage points more likely on a day with snowfall.

Next, I show that a directive nudge can affect students choices and help them improve their achievement. These results are more important than the weather IVs because they can be replicated in all types of classrooms. Students randomly assigned to the treatment group received an email in which I provide them with information about the negative correlation between procrastination and achievement. These students were 17% more likely (relative to a very low base rate) to successfully complete the course than students in the control group. Additionally, I show that the effect of the treatment is heterogeneous among different countries. For example, Germans assigned to the treatment group were 167% more likely to obtain the course certificate, Spaniards 67%, and Indians 40%.

The remainder of this paper is organized as follows. In the next section, I discuss the relevant literature. In Section 3.3, I present the economic model. Section 3.4, describes the data. In Section 3.5, I show that weather can be use as an instrument. Section 3.6, describes the randomized control trial and its results. In Section 3.7, I conclude.

3.2 Literature

Psychologists have been studying procrastination since the seventies. Ellis and Knaus (1977) claim that, “Procrastination constitutes an emotional hang-up that does you considerable damage.” They also claim that, based on their work as psychotherapists, about ninety-five percent of college-level individuals procrastinate. They never consider that they are basing their “guesstimate” from a highly selected sample (i.e., their patients). Neither did they consider that procrastination could be correlated with some other unobserved characteristic which is the real cause for their patients problems. Knaus (2001) describe procrastination as our “ancient nemesis.” He claims procrastination may have originated as early as 2.5 million years ago when our ancestors first grouped into small clans and someone decided to needlessly put off doing something beneficial for the clan. These hypotheses are founded with small surveys that rely on indirect measures of procrastination. Moreover, none of these studies addresses the problem of procrastination being an endogenous choice. Chun Chu and Choi (2005) are the first to consider that “active” procrastination could be good: some people choose to procrastinate because they know they will do better under pressure. For their study, they invited students to respond to a questionnaire entitled “Survey of University Students‘ Time Use.” 230 undergraduate students filled out the questionnaire, but the paper does not mention how these students compared to their peers who chose not to participate in the study. This raises concerns about selection bias and external validity. In this paper, I use a direct measure of procrastination and both an instrumental variable and experimental approach to deal with the endogeneity concerns.

The first paper to address procrastination in the economic literature is Akerlof (1991). Akerlof argues that although procrastination might initially appear to be outside the appropriate scope of economics, it affects the performance of individuals and institutions in the economics and social domain. He proposes an economic model in which procrastination occurs when present costs are unduly salient in comparison with future costs, leading individuals to postpone tasks until tomorrow without foreseeing that when tomorrow comes, the required action will be delayed yet again. This model challenges the common assumption in economics that individuals are rational maximisers. Anderson and Block (1995) argues that the examples Akerlof offers can be explained within the framework of the standard economic model. They argue that Akerlof confuses later regret with prior irrationality, which is parallel to confusing ex post with ex ante. O'Donoghue and Rabin (2001) develop a model where a person chooses from a menu of options and is partially aware of her self-control problems. Their model predicts that additional options can induce procrastination, and a person may procrastinate more in pursuing important goals than unimportant ones. They argue that their second result arises because the greater the effort a person intends to incur, the more likely she is to procrastinate in executing those plans. Instead of using the standard economic assumption that preferences are time-consistent (i.e., a person's relative preference for well-being at an earlier date over a later date is the same no matter when she is asked), they model individuals with present-biased preferences. Finally, Siegfried (2001) argues that combating procrastination is essential in order to have a successful undergraduate economics honors program. He argues that getting students to work on their thesis early is the key to success. In order to achieve this his university uses a series of short-term deadlines and the "fear of personal embarrassment." Banerjee and Duflo (2014) use data from edX to show a discontinuity on grades and "on time" enrollment. They argue that this suggest that "disorganization" is negatively correlated with performance.

One approach that I use to control for the endogeneity of procrastination is to use weather as an instrument. Connolly (2008) links the American Time Use Survey to rain data from the National Climatic Data Center (NCDC). She finds that, on rainy days, men shift on average 30 minutes from leisure to work, suggesting that rain raises the marginal value of work. In this paper, I link data from the NCDC on rain and snowfall to data from Coursera. Coursera collects student's IP addresses. Using this, I geolocate students and assign them a NCDC weather station. Assuming that procrastinators do not choose their location in response to the weather, rain and snow are an exogenous source of variation for procrastination and allows me to identify the causal effect of procrastination on achievement.

Martinez and Diver (2014) explore the opportunities and challenges that MOOCs generate for research. Using data from Coursera, they show a strong negative correlation between procrastination and achievement. In this paper, I go beyond studying the correlation to determine the causal effect of procrastination on achievement by using both an instrument variables and experimental approach. The experimental approach consists of nudging students by sending information in a email, as in Martinez (2014a). The term "nudge" was first used by Thaler and Sunstein (2008) to describe "Any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any option or significantly changing their economic incentives."

3.3 Model

In this section, I present a simple model in which a student with ability a chooses whether to take a Quiz in the first period, procrastinate and take it in the second, or not take it at all.[1] In the second period, if the student's best grade is greater than $\overline{g}$, the student gets a payoff equal to W. Not attempting the Quiz in a given period yields a grade equal to zero. On the other hand, attempting the Quiz yields an uncertain grade determined by a function of his ability and an unobserved random variable ϵ_t:

$$g_t = f\left(a, \epsilon_t\right).$$

Each period the student gets utility from non-Coursera activities (i.e., leisure, and work), ℓ_t. The marginal utility of these activities is given by μ_t. If students like to procrastinate, $\mu_1 > \mu_2$.

The utility of attempting the Quiz in period 2 for a student with a grade from period 1 equal to g_1 and ability a is given by the utility from non-Coursera activities when attempting the quiz, plus the expected payoff of succeeding in the course:

$$V_2^T\left(g_1, a\right) = \mu_2 \ell_{2,T} + \Pr\left[\max\left(g_1, g_2\right) > \overline{g} \middle| \, a\right] W.$$

The utility of not attempting the quiz in period 2 for a student with grade from period 1 equal to g_1 and ability a is given by the utility from non-Coursera activities when not attempting the quiz, plus the payoff of succeeding in the course if the grade in

[1] In this model, the student is rational and has time consistent preferences. The student will choose to procrastinate if that is the optimal choice given his information set. An alternative model could generate procrastination by assuming agents with time inconsistent preferences, as in Choi et al. (2003).

period 1 was greater than the threshold:

$$V_2^{NT}(g_1) = \mu_2 \ell_{2,N} + \mathbb{1}_{(g_1 > \overline{g})} W.$$

Therefore, the period 2 problem can be written as:

$$V_2(g_1, a) = \max\left\{V_2^T, V_2^{NT}\right\}.$$

The period 1 problem, in which the student decides whether to attempt the quiz or procrastinate, can be written as:

$$V_1(a) = \max\left\{\mu_1 \ell_{1,N} + E\left[V_2(0, a)\right], \mu_1 \ell_{1,T} + E\left[V_2(g_1, a)\right]\right\},$$

where grades are a function of student ability and an unexpected shock, and the utility from non-Coursera activities is greater when non attempting the quiz:

$$g_t = f(a, \epsilon_t)$$

$$\ell_{t,N} > \ell_{t,T}$$

This problem can be easily solved recursively. A student with $g_1 > \overline{g}$ will not attempt the quiz in period 2.[2] A student with $g_1 < \overline{g}$ will attempt the quiz if

$$\Pr\left[g_2 > \overline{g}\right] W > \mu_2\left(\ell_{2,N} - \ell_{2,T}\right)$$

[2]Adding direct utility from grades would allow this model to explain students attempting the quiz eventhough they do not need to do it in order to obtain W.

In the first period, a student will attempt the quiz if

$$\mu_1 \ell_{1,T} + E\left[V_2\left(g_1, a\right)\right] > \mu_1 \ell_{1,N} + E\left[V_2\left(0, a\right)\right]$$

$$E\left[V_2\left(g_1, a\right)\right] - E\left[V_2\left(0, a\right)\right] > \mu_1\left(\ell_{1,N} - \mu_1 \ell_{1,T}\right)$$

If the expected value of attempting the Quiz in period 1 increases, $\uparrow E\left[V_2\left(g_1, a\right)\right]$, or the expected value of going into period 2 with a grade of zero decreases, $\downarrow E\left[V_2\left(0, a\right)\right]$, then students who were at the margin of attempting the Quiz in period 1 will take it, that is, they will procrastinate less. We can interpret the empirical strategies which I employ in terms of changes in key parameters in the model.

The intention of the informational experiment described in section 6 is to increase the expected value of attempting quizzes earlier. Similarly, if the marginal utility of non-Coursera activities decreases, $\downarrow \mu_1$, students at the margin will procrastinate less. In section 5, I show that rain and snow affect procrastination.

3.4 Data

I use data from Coursera, and the National Climatic Data Center (NCDC) of the National Oceanic and Atmospheric Administration (NOAA).[3] Coursera data provides me with a measure of procrastination, and a platform to run an experiment. I link these data with NCDC data which provides weather conditions (rain and snow) for each Coursera participant in the geography identified by IP addresses.

[3]This weather data is publicly available at `http://goo.gl/x2UrXq`.

3.4.1 Coursera

I use data from the third edition of *Foundations of Business Strategy* (FBS) by Michael J. Lenox of the University of Virginia. 75,180 students were initially enrolled in this course. The course ran from January 13, 2014 to February 25, 2014. FBS. explores the underlying theory and frameworks that provide the foundations of a successful business strategy. The class is divided into weekly modules. Each module consists of an introductory video, a reading from the strategist's toolkit, a series of video lectures, a quiz, and a case study to illustrate points in the lectures. Students wishing to receive a Statement of Accomplishment must satisfy the following criteria:

1. Complete the 6 quizzes. Students can take each quiz three times. The recorded score is the best score on each quiz. Quizzes have 10 questions with 4 choices each.

2. Submit a final project: a strategic analysis for an organization of their choosing.

3. Assess five peers strategic analysis using the peer assignment function.

Final grades are out of 100 points. Out of the 100 points of the final grade, 50 points are for the final project, 42 points for the quizzes (spread evenly over the 6 quizzes), and 8 points for post and comment up-votes. Those who receive 70 points or more and assess five peers strategic analysis receive a Statement of Accomplishment.

In this paper I use the 24,122 students who expressed, at the time of enrollment, their intention to complete all the course work necessary to obtain the Statement of Accomplishment.[4] These students were randomly assigned to treatment and control

[4]At the time of enrollment students had to answer the following question: *"How many of the assignments and quizzes do you intend to do?"* Only students who selected all were considered for this study. Other students may have little or no intention of taking the quizzes.

groups to test whether a directive nudge affects their behavior and achievement. Only 1,212 of these students received a Statement of Accomplishment.

Figure 3.2 shows when each of the quizzes were published, and the timing of the directive nudge intervention, which occurred on the same day that Quiz 6 was published.

3.4.2 Weather Data

The NCDC has data from more than 90,000 weather stations around the world.[5] The data includes maximum and minimum daily temperatures, rain, and snowfall. Using stations and students, latitude and longitude, I can match each student to their nearest weather station.

Connolly (2008) shows that, for men, a rainy day shifts about half an hour from leisure and home production to work, suggesting that rain raises the marginal value of work.[6] As in her paper, I define a rainy day as a day with at least 0.10 inches of rain. Additionally, I take snow into consideration, and I define a snowy day as a day with at least 0.10 inches of snow. Rain and snow might have different effects on student choices. For example, in a rainy day a student may stay at work until later and be less likely to do Coursera work when he gets home. On the other hand, on a snowy day a student may have to stay home instead of going to work, increasing the likelihood of doing Coursera work.

[5]In this paper I use their daily data, but it is possible to access hourly records

[6]Result for women are weak. A rainy day is associated with 3 more minutes at work.

3.5 The Effect of Procrastination Using Weather as an Instrument

I estimate linear regressions using rain and snow as IVs for procrastination. The course started on Monday January 13, and the first quiz was published that day, (Figure 3.2). The OLS estimate in the first column of Table 3.1 shows that a student who takes the Quiz on day 1 is 15.4 percentage points more likely to obtain the statement of accomplishment than a student who does not, relative to a very low base of 0.6% who obtain the statement.

If procrastination is correlated with some other unobserved characteristic (e.g., ability), this estimate would be biased, probably upward. To deal with this endogeneity bias, I estimate a first stage relationship that shows that a student is 2.3 percentage points less likely to take the Quiz on the first day if it is raining, and 5 percentage points more likely if it is snowing. Connolly (2008) suggests that on a rainy day people are more likely to spend more time at work. If this is so, they would have less time to do their Coursera activities. On the other hand, on a snowy day people are more likely to stay home and do their Coursera work.

The second stage regression shows that being induced not to procrastinate in take the Quiz on day one increases the probability of obtaining the statement of accomplishment by 13.8 percentage points. Thus, the estimate shrinks because some good students self-select.[7]

[7]Note that Table 3.1 uses a linear probability model. Wooldridge (2012) says "...the linear probability model is useful and often applied in economics. It usually works well for values of the independent variables that are near the average in the sample". Angrist and Pischke (2008) give several empirical examples where the marginal effects of a dummy variable estimated by LPM and probit are indistinguishable. IV Probit can not be used here because the endogenous regressor is discrete. Vytlacil and Yildiz (2007) propose nonparametric identification and estimation of the

I can further investigate the impact of procrastination by consider how long it takes a student to attempt the quiz, rather than simply looking at whether they take it on the first day. However, I cannot observe this measure of procrastination for students who never took the Quiz. In order to have another estimate of the effects of procrastination on achievement I define an upper and lower bound for those students who did not take the Quiz. In Table 3.2, I assume that students who did not take the Quiz would have procrastinated until one minute after the deadline. In Table 3.3, I assume that students would have procrastinated for a year after the deadline. The IV estimate using the lower bound implies that an additional hour (week) of procrastination decreases the probability of obtaining the certificate by 0.01 (1.68) percentage points. The upper bound assumption, implies that a week of procrastination decreases the probability of obtaining the certificate by 0.336 percentage points. That is, procrastinating the day the Quiz is published has much larger effects than procrastinating an additional day after 10 days of procrastination.

3.6 Randomized Email Nudge

I divided up the group of 24,122 students into randomly assigned treatment and control groups. On February 17, 2014, the day Quiz 6 was published, I sent an email to the students in the treatment group, as shown in Figure 3.3. This email is a low-cost directive nudge encouraging students not to procrastinate.[8]

average effect of a dummy endogenous regressor in models where the regressors are weakly but not additively separable from the error term. A dynamic discrete choice model with 41 periods and new choices appearing every week could be a better specification. But that specification is more costly to solve and does not provide any clear advantage over this simple specification.

[8] I can confirm that 43.02% of these emails were open. To know this, a unique white pixel is inserted in the body of the email. When the white pixel loaded from the server, I know the email was open. Because images are blocked by default on some email clients, these numbers are lower

Table 3.4 shows the effect of the treatment on Quiz 6 outcomes. In order to deal with attrition, I assign a grade of 0 to students who did not attempt the quiz. "Took Q6" shows that students in the treatment group were 0.8 percentage points, or 7.74%, more likely to take the quiz. "maxGrade" shows that students assigned to the treatment group scored 7 points, or 8.92% better on their best attempt on Quiz 6 than students in the control group. "Procrastination" suggests that, conditional on taking the quiz at some point, students assigned to the treatment group procrastinated 2.1 fewer hours on average. This difference is not statistically significant.

The nudge not only affected students' behavior in Quiz 6, but also their choices to go back and take quizzes 1 to 5 for the first time. For example, Table 3.5 shows that students receiving the email nudge were about a half percentage point more likely to attempt Quizzes 1-5 for the first time after the intervention than students in the control group. The students who were nudged into taking the quizzes are increasing my measure of procrastination in the treatment group.[9]

Students assigned to the treatment group were not only more likely to complete Quiz 6, but also to obtain a Statement of Accomplishment. Table 3.6 shows that students assigned to the treatment group were 0.8 percentage points, or 16.85% more likely to obtain the course certificate. For an intervention with negligible cost, this is an important effect.

Table 3.7 shows that the treatment has heterogeneous effects across countries. Italians assigned to the treatment group procrastinated, on average, 63.22 fewer hours

bounds for the open rates. The open rate is crucial to calculating the effect of the treatment on the treated.

[9]For example, imagine that there are only 2 students in the treatment and 2 in the control group. One student in each group took Quiz 1 before the intervention. After the intervention, the student in the treatment group decides to stop procrastinating and take the Quiz, but the student in the control group keeps on procrastinating. This will increase my measure of procrastination for the treatment group.

than those assigned to the control group. Nigerians and Indians assigned to the treatment group procrastinated more than those in the control group. This is explained by the fact that those assigned to the treatment group still procrastinated but were more likely to take the quiz at all.

Finally, Table 3.8 shows that the effect on outcomes is also heterogeneous across countries. Germans assigned to the treatment group were 167% more likely to obtain the certificate, Spaniards 67%, Indians 40%, and there were not statistically significant effects for other countries.

3.7 Conclusions

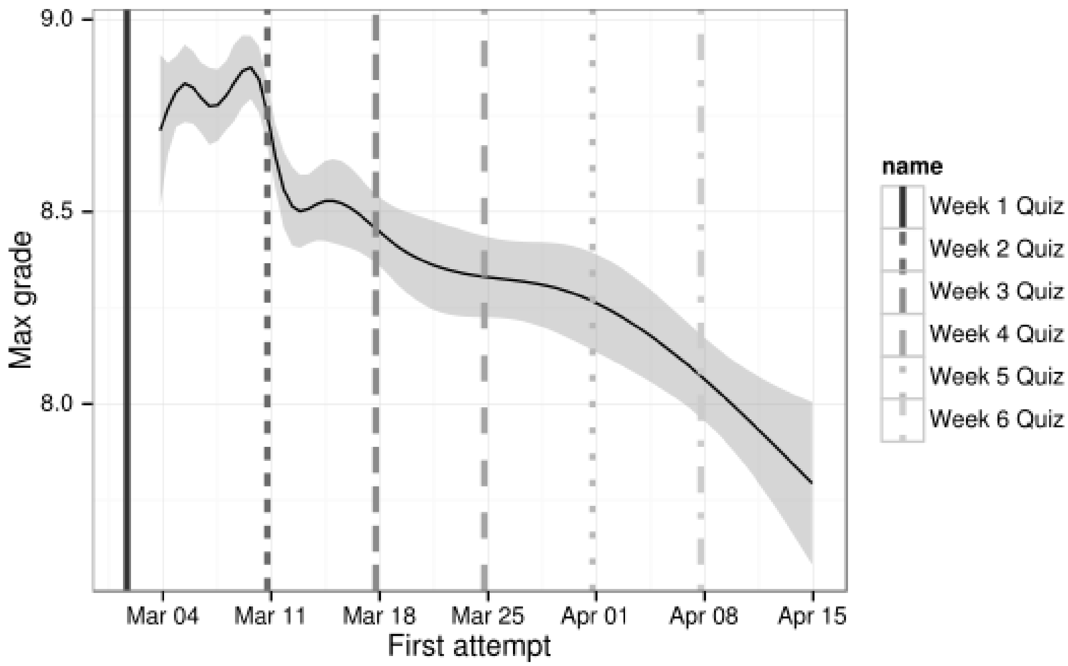
This paper examines the role of procrastination on achievement. Understanding whether or not there is a causal relationship, so that inducing procrastinators to take action improves their learning, is fundamental in order to design a course with incentives that maximize student achievement.

First, I show that a student is less likely to take a quiz on a rainy day and more likely to take a quiz on a snowy day. Using weather as an instrument, I show that attempting the first quiz on the day it is published increases the probability of obtaining the Statement of Accomplishment by 13.8 percentage points.

Second, I show that a directive nudge can strongly increase achievement. For example, students in the treatment group were 0.8 percentage points, or 16.85% more likely to obtain the course certificate. For an intervention with negligible cost, this is an important effect. Moreover, the effects are heterogeneous across countries. For example, Germans assigned to the treatment group were 167% more likely to obtain the course certificate, Spaniards 67%, and Indians 40%. There were no statistically significant effects in achievement for students from other parts of the world, especially as sample sizes got smaller. In order to understand the causes of this heterogeneity, more data is needed. For example, it is possible that the level of education are different, or employment status, or simply cultural differences. Currently, Coursera survey data is not good enough to address this question, but Coursera is constantly improving their platform.

Future research should explore other directive nudges to improve achievement. For example, evidence Martinez (2014a) shows that telling students how they are performing relative to their peers can improve ranking, on average, by 8.43 percentage points. Future intervention could test what is the impact of telling students how students in the top of the class are using their time before attempting a quiz. Additionally, there are also the questions of whether there is a critical time for nudging, and whether follow-ups help.

Figure 3.1: Procrastination and Achievement



Note: Extracted from Martinez and Diver (2014).

Figure 3.2: Time Line

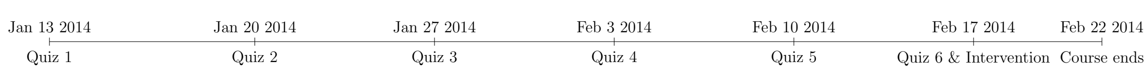
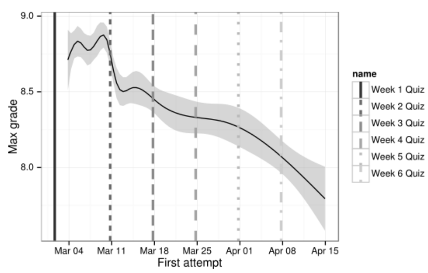


Figure 3.3: Treatment

Subject: [Foundations of Business Strategy] Don't Leave for Tomorrow What You Can Do Today

Dear [name],

Our analysis of the previous iteration of this course shows that students who choose to do the quizzes late perform worse than those who do them earlier.



We encourage you to try the next quiz earlier. Keep in mind that you can retake the quizzes 3 times.

Best,

University of Virginia MOOC Research Team

Table 3.1: Regression Results, taking Quiz 1 the day is published and achievement

	Dependent variable:		
	Certificate	took Quiz 1 the 13	Certificate
	OLS	*First Stage*	*Second Stage*
	(1)	(2)	(3)
took Quiz 1 the 13	0.154***		0.138**
	(0.003)		(0.063)
rain the 13[a]		−0.023***	
		(0.006)	
snow the 13[b]		0.050**	
		(0.020)	
longitude		−0.0002***	
		(0.00004)	
Constant	0.006***	0.291***	0.011
	(0.002)	(0.004)	(0.018)
Observations	23,463	23,463	23,463
R^2	0.102	0.002	0.101
Adjusted R^2	0.102	0.002	0.101
Residual Std. Error	0.206 (df = 23461)	0.451 (df = 23459)	0.206 (df = 23461)
F Statistic	2,678.940*** (df = 1; 23461)	17.446*** (df = 3; 23459)	

Notes: *p<0.1; **p<0.05; ***p<0.01
[a] dummy variable equal to 1 if it rained more than 0.1 inches the day Quiz 1 was published.
[b] dummy variable equal to 1 if it snowed more than 0.1 inches the day Quiz 1 was published.

Table 3.2: Regression Results, Procrastination Q1 and Achivement

	Dependent variable:		
	Certificate	ProcrastinationLow	Certificate
	OLS	*First Stage*	*Second Stage*
	(1)	(2)	(3)
ProcrastinationLow[a]	-0.0002^{***}		-0.0001^{**}
	(0.00000)		(0.0001)
rain the 13[b]		18.949^{***}	
		(5.243)	
snow the 13[c]		-46.341^{***}	
		(17.248)	
longitude		0.220^{***}	
		(0.034)	
Constant	0.196^{***}	816.126^{***}	0.160^{***}
	(0.003)	(3.266)	(0.055)
Observations	23,463	23,463	23,463
R^2	0.102	0.003	0.096
Adjusted R^2	0.102	0.003	0.096
Residual Std. Error	0.206 (df = 23461)	389.703 (df = 23459)	0.207 (df = 23461)
F Statistic	$2{,}676.202^{***}$ (df = 1; 23461)	20.741^{***} (df = 3; 23459)	

Notes: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

[a] how many hours pass between the publishing of the Quiz and their first attempt at it. For students who did not attempt the Quiz, I assume they procrastinated until a minute after the deadline.

[b] dummy variable equal to 1 if it rained more than 0.1 inches the day Quiz 1 was published.

[c] dummy variable equal to 1 if it snowed more than 0.1 inches the day Quiz 1 was published.

Table 3.3: Regression Results, Procrastination Q1 and Achivement

	Dependent variable:		
	Certificate	ProcrastinationHigh	Certificate
	OLS	*First Stage*	*Second Stage*
	(1)	(2)	(3)
ProcrastinationHigh[a]	-0.00002^{***}		-0.00002^{**}
	(0.00000)		(0.00001)
rain the 13[b]		243.716^{***}	
		(57.891)	
snow the 13[c]		-457.968^{**}	
		(190.432)	
longitude		2.023^{***}	
		(0.374)	
Constant	0.167^{***}	$6,673.100^{***}$	0.153^{***}
	(0.002)	(36.055)	(0.044)
Observations	23,463	23,463	23,463
R^2	0.118	0.002	0.116
Adjusted R^2	0.118	0.002	0.116
Residual Std. Error	0.204 (df = 23461)	4,302.763 (df = 23459)	0.204 (df = 23461)
F Statistic	$3,138.606^{***}$ (df = 1; 23461)	17.307^{***} (df = 3; 23459)	

Notes: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

[a] how many hours pass between the publishing of the Quiz and their first attempt at it. For students who did not attempt the Quiz, I assume they procrastinated for a year.

[b] dummy variable equal to 1 if it rained more than 0.1 inches the day Quiz 1 was published.

[c] dummy variable equal to 1 if it snowed more than 0.1 inches the day Quiz 1 was published.

Table 3.4: Treatment effect on Quiz 6 outcomes and choices

	Treatment	Control	90% Confidence Interval	p-value
maxGrade[a]	0.8506	0.7809	(0.0154, 0.124)	0.0347
firstGrade[b]	0.5278	0.4966	(-0.005, 0.0673)	0.1561
Attempts[c]	0.2461	0.2241	(0.0059, 0.0381)	0.0247
Took Q6[d]	0.1002	0.093	(0.001, 0.0135)	0.058
Procrastination[e]	102.9237	105.0334	(-6.8086, 2.5892)	0.4601
n	12061	12061		

Sample: Students who stated their intention to complete all the work needed to obtain the certificate
[a] First attempt grade
[b] Best attempt grade
[c] Numbers of attempts
[d] Dummy equal to 1 if the student attempted the Quiz
[e] Number of hours between the quiz publication and student first attempt

Table 3.5: Attempting Quizzes 1-5

	Treatment	Control	90% Confidence Interval	p-value
Quiz 1	0.0198	0.017	(0, 0.0057)	0.1034
Quiz 2	0.0273	0.024	(0, 0.0067)	0.1031
Quiz 3	0.0343	0.0296	(0.001, 0.0085)	0.0369
Quiz 4	0.0424	0.0383	(-1e-04, 0.0082)	0.1088
Quiz 5	0.0585	0.0522	(0.0015, 0.0111)	0.0323
n	12061	12061		

Table 3.6: Achievement

	Treatment	Control	90% Confidence Interval	p-value
Certificate	0.0541	0.0463	(0.0032, 0.0124)	0.0056
n	12061	12061		

Table 3.7: Procrastination by Country

Country	treatment	control	90% Confidence Interval	p-value	n Treatment	n Control
Italy	62.95	126.17	(-96.4735, -29.9799)	0.00	61	53
Nigeria	115.57	87.28	(7.753, 48.8257)	0.03	79	65
India	117.28	103.78	(1.7414, 25.258)	0.06	215	194
Mexico	112.27	150.03	(-70.4457, -5.074)	0.06	63	55
Netherlands	57.34	94.35	(-73.2042, -0.8067)	0.09	58	63
Portugal	113.85	66.98	(-1.1174, 94.861)	0.11	59	53
Canada	97.89	117.06	(-41.7352, 3.3982)	0.16	102	82
United States	99.96	107.54	(-17.3334, 2.1589)	0.20	316	314
Australia	97.61	81.27	(-11.7926, 44.4721)	0.33	74	59
Russia	89.41	114.47	(-68.7584, 18.6417)	0.34	58	62
United Kingdom	97.72	114.48	(-47.8758, 14.3557)	0.37	90	78
China	82.74	96.13	(-45.1692, 18.3865)	0.48	62	60
Spain	95.39	85.30	(-15.8501, 36.0327)	0.52	90	78
Colombia	88.31	100.23	(-58.5295, 34.6981)	0.66	56	57
France	104.11	97.24	(-19.9504, 33.6842)	0.67	76	70
Germany	100.89	95.85	(-26.5307, 36.6173)	0.79	69	54
Ukraine	82.34	87.62	(-47.6962, 37.1344)	0.83	58	54
Brazil	113.93	112.98	(-14.9745, 16.8918)	0.92	123	110

Table 3.8: Course completion

Country	treatment	control	90% Confidence Interval	p-value	n Treatment	n Control
India	0.07	0.05	(0.0041, 0.0338)	0.04	1749	1677
Germany	0.08	0.03	(0.01, 0.0954)	0.04	482	506
Spain	0.10	0.06	(0.004, 0.0788)	0.07	615	602
Portugal	0.06	0.03	(-0.0089, 0.0769)	0.19	455	455
Brazil	0.04	0.03	(-0.0027, 0.0233)	0.19	1334	1336
France	0.10	0.08	(-0.0242, 0.0754)	0.40	509	506
Nigeria	0.08	0.06	(-0.0211, 0.0601)	0.43	545	521
Netherlands	0.08	0.11	(-0.0889, 0.0362)	0.49	450	437
Canada	0.09	0.08	(-0.0206, 0.0443)	0.55	725	697
Russia	0.03	0.04	(-0.0378, 0.0177)	0.55	584	559
United Kingdom	0.05	0.04	(-0.0149, 0.0291)	0.60	784	827
Italy	0.07	0.06	(-0.0366, 0.0667)	0.63	453	440
Ukraine	0.04	0.03	(-0.0231, 0.0414)	0.64	508	485
Colombia	0.04	0.05	(-0.0574, 0.0322)	0.64	448	446
Mexico	0.03	0.02	(-0.0204, 0.0336)	0.69	499	538
Australia	0.06	0.05	(-0.0275, 0.0418)	0.73	598	528
China	0.03	0.03	(-0.0182, 0.0261)	0.77	627	623
United States	0.05	0.05	(-0.0088, 0.0096)	0.94	3226	3200

Chapter 4

MOOCs as a brick-and-mortar complement.

Coauthored with Louis Bloomfield and Sarah Turner.

4.1 Introduction

The gap between high- and low-income families in college entry, persistence and graduation is growing, as shown by Bailey and Dynarski (2011) using 70 years of data. Part of this graduation gap might be explained by differences in the ability to pay private tutors. The private tutor market is expected to surpass $102.8 billion by 2018 Crotty (2012). A world with high-priced tutors raises concerns about fairness: wealthy students can afford them while poor ones cannot. A tutor in Manhattan can charge up to $400 an hour, while the online tutoring website TutorVista.com charges $45 for two hours of tutoring Sullivan (2010). There is no study that compares the effectiveness of traditional tutors versus more affordable online tutoring.

Can we improve student achievement in a brick-and-mortar classroom by encouraging participation in a Massive Online Open Course (MOOC)? Because enrollment in the MOOC is endogenous, we cannot simply compare the mean grade in the brick-and-mortar course for students who enroll or not in the MOOC. For example, low-ability students may be more likely to enroll in the MOOC than high-ability students because they need all the help they can get. Students enrolled in Professor Bloomfield's *How Things Work* (PHYS1060) in Spring 2014 at the University of Virginia were randomly assigned to one of two treatments, or a control group. The 91 students in treatment 1 received an email offering them $10 if they enrolled in the Coursera version of the course, as shown in Figure 4.1. The 91 students in treatment 2 received an email offering them $50 if they obtain 80% in the Coursera version of the course, as shown in Figure 4.2.

The Coursera version of the course covers the same material as the first six weeks of the bricks-and-mortar version. Coursera students are able to pause or fast-forward Professor Bloomfield's videos, make him talk faster or slower, or make him repeat the same thing a thousand times. Additionally, Coursera allows students to take the quizzes as many times as they want, and provides them with a forum in which they can ask questions anonymously.[1]

The existing literature on conditional transfers designed to encourage students to exert more effort towards their studies finds mixed, and often small, impacts on achievement. Despite a low take up rate for the experiment, we find that the treatments increase the probability that students enroll in Coursera, and that enrolling in Coursera causes higher achievement. Although our estimates are not very precisely estimated due to a small sample size, our findings suggest a much greater impact per dollar than those in the existent literature.

The remainder of this paper is organized as follows. In the next section, we discuss related research. In Section 4.3 we present an economic model. Section 4.4 describes the data. Section 4.5 presents the results. In Section 4.6 we conclude.

[1] The quizzes in the online and brick-and-mortar versions of the course were not the same.

4.2 Related Literature

The literature on conditional transfers to encourage students to exert more effort towards their studies finds mixed, and often small, impacts on achievement. Jackson (2010) shows that a program implemented in Texas that pays both teachers and students for passing grades on advanced on Advanced Placement (AP) examinations increases AP course and exam taking, the number of students with high SAT/ACT scores, and college matriculation. This program costs about $200 per student taking an AP exam. On the other hand, Fryer (2011) finds that the impact of financial incentives on student achievement is statistically indistinguishable from 0. He paid second graders in Dallas $2 per book read if they passed a short quiz to confirm they had read it; fourth- and seventh-grade students from NYC were paid for performance on a series of ten interim assessments; and ninth graders from Chicago were paid every five weeks for grades in five core courses. In post-secondary education, Angrist, Lang, and Oreopoulos (2009) evaluate The Student Achievement and Retention (STAR) project. This program provided different support and financial incentives to help improve academic performance in college. These are programs that cost between $300 and $740 per student. They conclude that the cost of these will be more than offset by future earnings gains for women but that the programs were ineffective for men. Angrist, Oreopoulos, and Williams (2010) try to replicate the results from STAR with a new program, the "Opportunity Knocks" (OK). OK awards were more generous: the expected OK award was $1,330 while the expected STAR award was about $400. While STAR included the opportunity to participate in study groups, OK services consisted of email-based peer mentoring. The OK program failed to replicate the strong positive results for women seen in STAR. Barrow and Rouse (2013) find that students eligible for a performance-based scholarship devoted more time to educational activities. They show that these changes in behavior do not persist beyond eligibility for the scholarship suggesting that incentives do not permanently change their cost of effort or their ability to transform effort into educational outcomes. They also find that larger payments did not generate larger increases in effort. The fact is that financial incentives have small effects on academic outcomes. Barrow et al. (2012) evaluate the effect of performance-based incentive programs on educational outcomes for community college students. The cost of this program was approximately $1,100 per student. They find that students in the treatment group earned 3.69 credits more than control group students. This translates into an additional 1.23 courses, which they estimate translates into $123 per year in annual earnings. They further assume this value stays constant in real terms, and that over 20 years 1.23 courses is equivalent to $2,977. Summarizing, these studies find, at best, that the impact of monetary incentives on outcomes are small.

Other studies have focused on non-financial incentives to help students improve their performance. Trost and Salehi-Isfahani (2012) assess the effect of the completion of online homework assignments on exam performance in "Principles of Economics." They find that completing homework assignments early in the course has a modest effect on related questions on the midterm exam. However, the effect of missing one homework assignment does not negatively affect final exam performance. In Martinez (2014a), I evaluate the impact of providing students with information about their performance relative to their classmates. I find evidence that students respond to this informational nudge and that framing matters. Students who were doing relatively poorly respond to the negative treatment with more effort, and this effort translates, in some cases, into higher achievement. On the other hand, students who were doing relatively well respond to the positive treatment. As an example of the magnitude of the effects, the average student in the control group, who before the intervention did not have a perfect score in the first quiz, was ranked in the 31.6 percentile of the class in the third quiz, while the average student in the negatively framed treatment was in the 40.5 percentile. Additionally, in Martinez (2014b), I show that a directive nudge can decrease students' procrastination and increase their achievement.

To sum up, interventions that cost between \$200 and \$1,000 per student have very small effects on achievement and are unlikely to be scalable. In this paper, we study an intervention that has a very low marginal cost and is easily scalable.

4.3 Theoretical Framework

We write down a model of student investment in education that shows that reducing the cost of exerting effort increases achievement. As in Becker (1967), students invest in their education until the marginal cost of doing so equals the marginal benefit. Suppose that student i's grade, g_i, depends on ability a_i, effort e_i, and some random noise ϵ_i as follows:

$$g_i = \alpha_0 + \alpha_1 e_i + \alpha_2 a_i + \epsilon_i,$$

where α_0, α_1, and α_2 are all positive parameters of the grade production function. Let ϵ be distributed $F(\epsilon)$ with density $f(\epsilon)$, and let $c(e)$ reflect the cost of effort. Assume $c(e)$ is an increasing, concave, and twice differentiable function. Further assume there is a payoff W for achieving at least a minimum grade, $g^{\min}$, with a payoff of zero otherwise.

Assuming students maximize utility by maximizing the net expected benefit of effort, a student's maximization problem is as follows:

$$\max_{e} \left\{1 - F\left(g^{\min} - \alpha_0 - \alpha_1 e_i - \alpha_2 a_i\right) \cdot W - c\left(e\right)\right\}$$

subject to

$$e \geq 0$$

were $F(\cdot)$ is the probability that a student with ability a_i who exerts effort e_i obtains a grade lower than $g^{\min}$. That is, the student is choosing e to maximize the difference between the expected benefit of exerting effort e and the cost of exerting that effort.

The optimal value of effort, $e_i^{\star}$, is characterized by the following conditions:

$$\alpha_1 f\left(g^{\min} - \alpha_0 - \alpha_1 e_i - \alpha_2 a_i\right) \overline{W} \quad \leq \quad c'\left(e_i^{\star}\right),$$

$$e_i^{\star} \geq 0, \text{ and}$$

$$e_i^{\star}\left[\alpha_1 f\left(g^{\min} - \alpha_0 - \alpha_1 e_i^{\star} - \alpha_2 a_i\right) \overline{W} - c'\left(e_i^{\star}\right)\right] \quad = \quad 0.$$

If the marginal benefit is relatively low, reflected in, $f\left(g^{\min} - \alpha_0 - \alpha_1 e_i - \alpha_2 a_i\right)$, or the marginal costs are relatively high, reflected in $c'\left(e_i^{\star}\right)$, a student may not exert any effort. A decrease in the cost of exerting effort will lead to an increase in effort for students who are already exerting positive effort and some of those who are at the margin between exerting effort or not.

The availability of the Coursera version of the course reduces the cost of exerting effort in the brick-and-mortar course. Watching video lectures, and doing online exercises with instantaneous feedback is easier, and possibly more entertaining, for students than going to the library. If this is the case, we should expect students in the treatment group to enroll in Coursera and perform better in the course. An alternative interpretation, which we cannot distinguish, is that this intervention is increasing the marginal benefit of exerting effort, α_1. If this is the case, students in the treatment group would also perform better in the course.

4.4 Data

4.4.1 How Things Work

How Things Work (PHYS 1060) is an unconventional introduction to physics, a course that starts with whole objects and looks inside them to see what makes them work. Effectively "case-study physics," it is designed for non-science students who are looking for real-world relevance in their studies. Discussions with the instructor suggest that when physics is taught in the context of ordinary objects, these students are enthusiastic about it, look forward to classes, ask insightful questions, experiment on their own, and find themselves explaining to friends and family how things in their world work. The focus of PHYS 1060 is on simple mechanical objects; electromagnetic objects; objects involving radio waves, microwaves, and light; objects that use optics; and nuclear objects. This paper uses data from 234 students enrolled in PHYS 1060 in spring 2014 at the University of Virginia.

One of the first courses offered through a partnership between Coursera and the University of Virginia was a MOOC version of *How Things Work*, which was available for free to anyone around the world. Video lectures lie at the heart of MOOCs: these are the equivalent of going to a lecture. The forums are potentially a very important tool in a MOOC. In a traditional classroom, students can go in during office hours or talk to classmates if they are having trouble with a concept; the forums play that role in MOOCs. Finally, the online quizzes are the assessments used to measure mastery of class material. The Coursera MOOC covers the Laws of Motion and lays the groundwork for many subsequent concepts in PHYS 1060. A student who does not understand the Laws of Motion will struggle with subsequent coursework in the semester. For example, Professor Bloomfield introduces the concept of energy in the MOOC component of PHYS 1060 and discusses energy frequently for the rest of the semester.

The 273 students initially enrolled in the course were divided into treatment 1 (the \$10 treatment), treatment 2 (the \$50 treatment), and the control group (C). Of these students, 33 did not take the first midterm test, 4 did not take the pre-test at the beginning of the semester, and 2 did not take the final exam. After removing these students, our sample size is 234 students: 78 were assigned to the control group, 74 were assigned to the \$10 treatment, and 82 to the \$50 treatment.[2] This course started on January 13, 2014 and ended on May 5, 2014. Figure 4.3 shows the course timeline.

Professor Bloomfield gave his students, both in the MOOC and brick-and-mortar courses, a pre-test to measure their initial understanding of physics. Figure 4.4 shows the grade distribution for the pre-test. This pre-test was available to students from January 4 to January 22 from the website they used for homework. On January

[2]If attrition is not random, for example, if low-ability students are more likely to drop out if assigned to the treatment group, then this would bias our estimates.

9, four days before the first class, a randomly selected group of students received an email nudging them to enroll in the Coursera version of the course. The first midterm test was on February 17. That day, students were surprised by having to retake the pre-test rather than new questions on the mid-term. Not surprisingly, Figure 4.5, shows a positive correlation between the pre-test grade and the midterm grade.

On the final exam, 11 out of 60 questions were specifically on material covered in the Coursera MOOC (the Laws of Motion). Figure 4.6 shows a positive correlation between the pre-test grades and final grades.

4.4.2 Follow-Up Survey

The response rate for the email nudges we sent on January 9 was very low. Only 16 out of 168 students emailed us back, 8 from the $10 treatment and 8 from the $50 treatment. To understand why they did not respond to our email, we surveyed the students on February 2. Out of the 273 students, only 168 completed the survey, so the response rate was higher than for the email treatments, but well less than 100%. Table 4.1 shows their responses. We asked students the following multiple choice questions:

1. In classes that offer optional (ungraded) materials online, such as videos and readings, what fraction of those materials do you use? [none; very few; some; most; all]

2. What grade do you expect to obtain in PHYS 1060? [A; B+; B; B−; C+; C; C−; less than C−]

3. What did you think about the email that encouraged you to enroll in the Cours-

era version of *How Things Work* [legitimate; a scam; didn't read; don't remember receiving; don't read email]

4. If you read that email, was the email clear? [Yes; No; Not sure]

5. How long do you think it would take you to sign up for the Coursera version of *How Things Work*? [Less than 2 minutes; Less than 10 minutes; More than 10 minutes; Not sure]

6. How long do you think it would take you to complete the Coursera version of *How Things Work*? [Less than 2 hours a week; Between 2 and 4 hours a week; More than 4 hours a week; Not sure]

7. Are you enrolled in the Coursera version of *How Things Work*? [Yes; No]

A second part of the survey indicated that it should only be answered by those who were not enrolled in the Coursera version of *How Things Work*. The questions were:

1. The reason you did not enroll is [You think it is not worth your time; The email you received was confusing and you did not know what to do; You did not know about the existence of the course; Other]

2. If you were to be paid to try an online course like Coursera's *How Things Work*, the minimum you would require is [$0; $5; $10; $25; $50; $75; $100; More than $100]

The main findings from the survey are that the email was not clear enough, and that $10 is too little to encourage students to enroll in Coursera. Only 37% thought the email was "legitimate." In the next iteration of this experiment, emails will be sent from the professor's account. In addition to use bigger monetary incentives, we plan to test informational nudges in which we tell students how they performed in the pre-test relative to other students and suggest them to try the MOOC. Fortunately, 32 students responded that they were enrolled in the Coursera version of the course. In Section 4.5, we estimate the causal effect of enrolling in Coursera using the emails as instruments for Coursera enrollment.

4.5 Estimation Results

Professor Bloomfield's students were asked to take a pre-test to measure their knowledge of physics. This pre-test consisted of questions from the previous year's exam. On February 17, the day of the midterm, students were surprised by having to take the pre-test again instead of new questions. To study how much the students had improved, we look at the difference between the pre-test and the midterm. Figure 4.7, shows the improvement distribution.

A naive approach to examining how Coursera enrollment affects student improvement would be to regress improvement in Coursera enrollment. The first column in Table 4.2 shows that, on average, students who enroll in Coursera improve their grades between the pre-test and the midterm by 6.6 points more than students who choose not to enroll. However, this estimate is likely to be biased. Students who choose to enroll in Coursera are those who probably need all the help they can get.[3] If that is the case, on average, a student who enrolls in Coursera is less likely to improve because of some unobservable characteristics.

[3] Alternatively, these students may have a lower cost of exerting effort.

To deal with the endogeneity of enrolling in Coursera, we use the experimental treatments as instruments for enrollment. Column 2 in Table 4.2 shows that students assigned to the \$10 treatment or the \$50 treatment are 12.4 and 13.2 percentage points more likely to enroll in Coursera, respectively, relative to a baseline enrollment rate of 5.1% within the control group. Column 3 shows that after controlling for endogeneity, the estimate of the effect of MOOC enrollment on the mid-term test scores goes from 6.6 to 14.2, though it is not statistically significant. Finally, column 4 shows that enrolling in Coursera causes an average improvement between the pre-test and the final of 68.9 points, relative to a baseline score of 19.4. That is, on average, if students enroll in Coursera they will improve their grades for the final by 255% more than if they do not. Because of our small sample size, our estimates are imprecise. The 90% confidence interval is from 0.82 to 98.15 points.

It is possible that students who performed really poorly on the pre-test cannot benefit from Coursera. To explore this hypothesis, we restrict the sample to the 75% of students who obtained more than 30 in the pre-test. Table 4.3 shows that students assigned to the \$10 treatment or the \$50 treatment are 17.1 and 11.5 percentage points more likely to enroll in Coursera, respectively. The instrumental variable results show that students enrolled in Coursera significantly improve their scores, on average, by 37.05 points more than students who choose not to enroll. The 90% confidence interval goes from 0.93 to 73.17. Moreover, enrolling in Coursera increases the improvement between the pre-test and the final by 61.14 points, with a 90% confidence interval from 9.28 to 111.00. That is, on average, if students enroll in Coursera they will improve their grades for the final by 600% more than if they do not. Therefore, we find supporting evidence of heterogeneous effects of Coursera on test outcomes by pre-course ability/knowledge.

4.6 Conclusions

The main contribution of this paper is to show that a MOOC can serve as a complement to a bricks-and-mortar course. Using small monetary incentives, \$10 or \$50, we encouraged a randomly selected group of students to take the MOOC version of the course. We show that enrolling in the MOOC causes students' grades to improve more compared to those who do not enroll. The cost for a student is time, and the benefits are an increase in performance and a decrease in tutoring. These benefits are probably bigger in other universities where dropout and failing are more frequent.

Although the results of this paper are promising, there are several limitations that we intend to address in future, related work. First, although we are able to identify a positive causal effect of enrolling in Coursera on bricks-and-mortar achievement, these point estimates are imprecise due to the small sample size and the low take-up rate. We intend to run a similar experiment with more students, including students from other universities. Second, the survey reveals that some students did not trust the source of the emails nudging them into enrolling in Coursera. In the next iteration of this experiment, emails will be sent from the professor's account. Finally, the economic incentives were relatively small. Many of the students who qualified to claim the gift card by signing up after receiving the treatment email did not claim it. In the future iteration of this research, one of the treatments will be a more significant economic incentive, and the other an informational nudge as in Martinez (2014a). Finally, out of the 32 students who told us in the survey that they were enrolled in Coursera, we were able to link only five to the Coursera data. This is because most students do not use their university email account to register in Coursera. Getting information about their Coursera login credentials will allows us to see how students interact with the MOOC.

Figure 4.1: Treatment 1

Subject: How Things Work

Dear [name],

Did you know that there is an online course available through Coursera that could help you succeed in Professor Bloomfield's class "How Things Work"? Signing up is easy – just go to `https://www.coursera.org/course/howthingswork1` and click the blue "Learn for Free" button. It will only take you few seconds.

Because we would like to encourage you to try the "How Things Work" material on Coursera, we will offer you a $10 Amazon gift card if you enroll in this Coursera course by January 20. To obtain the gift card, email mooc-research@virginia.edu with the email address you are using for the Coursera course (e.g., abc1yz@virginia.edu). If you find that the Coursera course is not useful, you can always stop participating without any consequence.

Ready to get started? Sign up at `https://www.coursera.org/course/howthingswork1` and then send an email to mooc-research@virginia.edu. If you have other questions, please do not hesitate to contact us at mooc-research@virginia.edu.

With all good wishes for the spring term,

University of Virginia MOOC Research Team

Figure 4.2: Treatment 2

Subject: How Things Work Dear [name],

Did you know that there is an online course available through Coursera that could help you succeed in Professor Bloomfield's class "How Things Work"? Signing up is easy – just go to `https://www.coursera.org/course/howthingswork1` and click the blue "Learn for Free" button. It will only take you few seconds.

Because we would like to encourage you to try the "How Things Work" material on Coursera, we will offer you a $50 Amazon gift card if you enroll in this Coursera course by January 20 and obtain a Coursera statement of accomplishment with a grade not lower than 80%. To obtain the gift card, email mooc-research@virginia.edu with the email address you are using for the Coursera course (e.g., abc1yz@virginia.edu). If you find that the Coursera course is not useful, you can always stop participating without any consequence.

Ready to get started? Sign up at `https://www.coursera.org/course/howthingswork1` and then send an email to mooc-research@virginia.edu. If you have other questions, please do not hesitate to contact us at mooc-research@virginia.edu.

With all good wishes for the spring term,

University of Virginia MOOC Research Team

Figure 4.3: Timeline

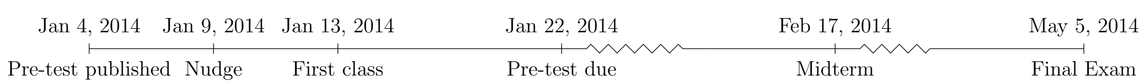


Figure 4.4: Pre-test grade distribution

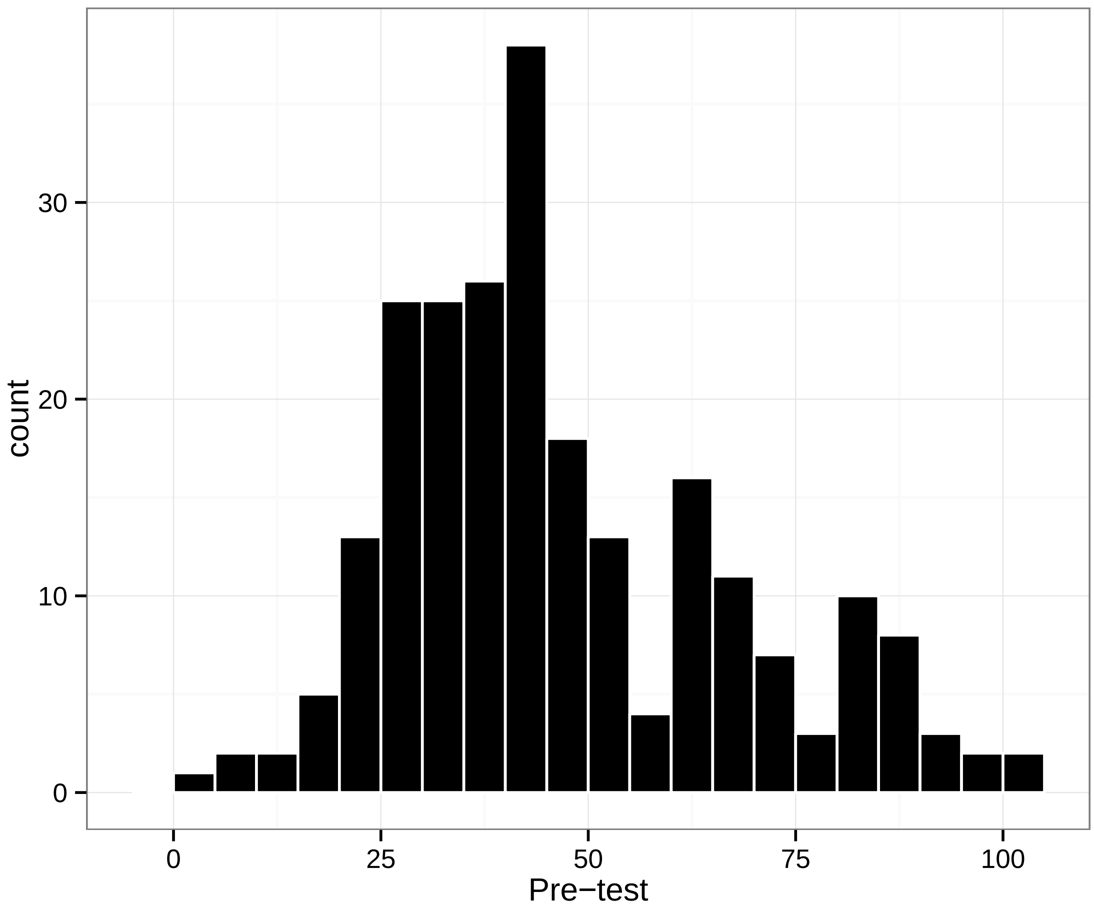


Figure 4.5: Midterm and pre-test grades

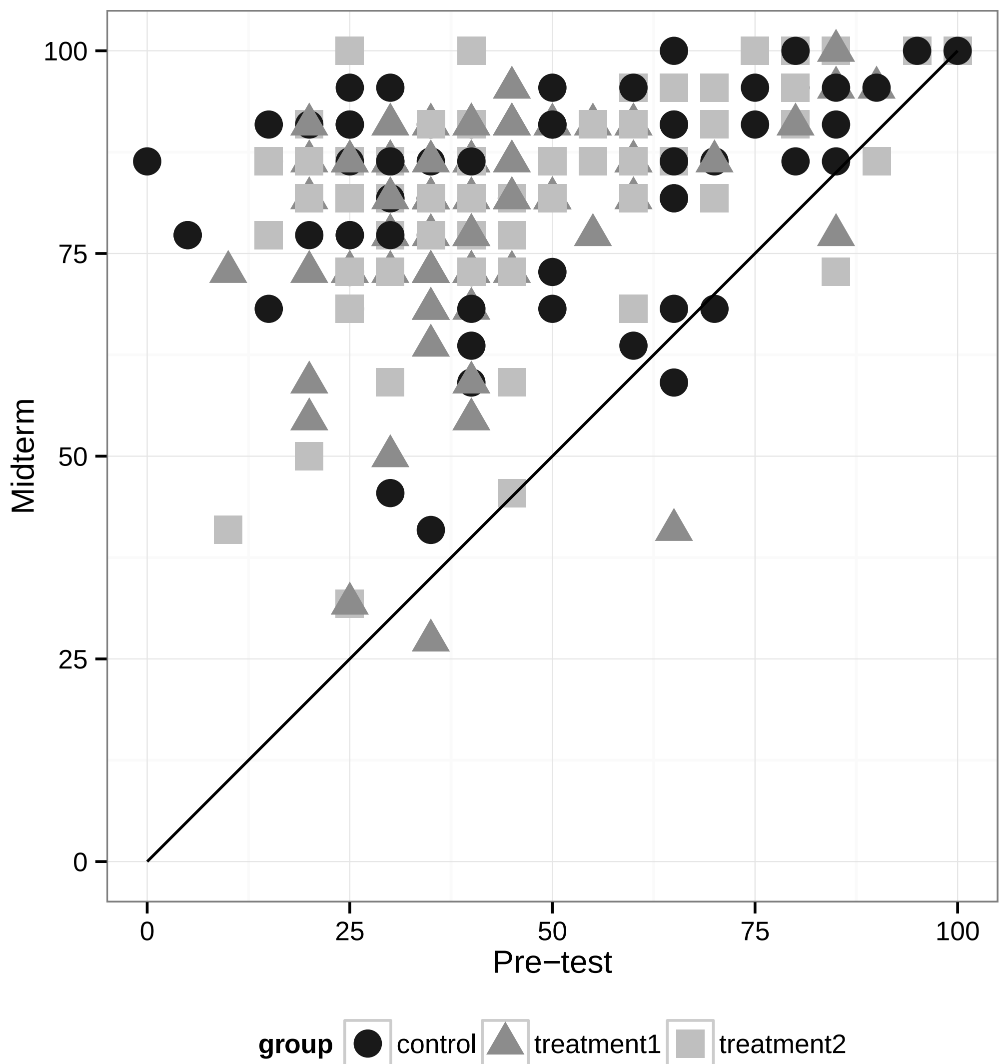


Figure 4.6: Final and pre-test grades

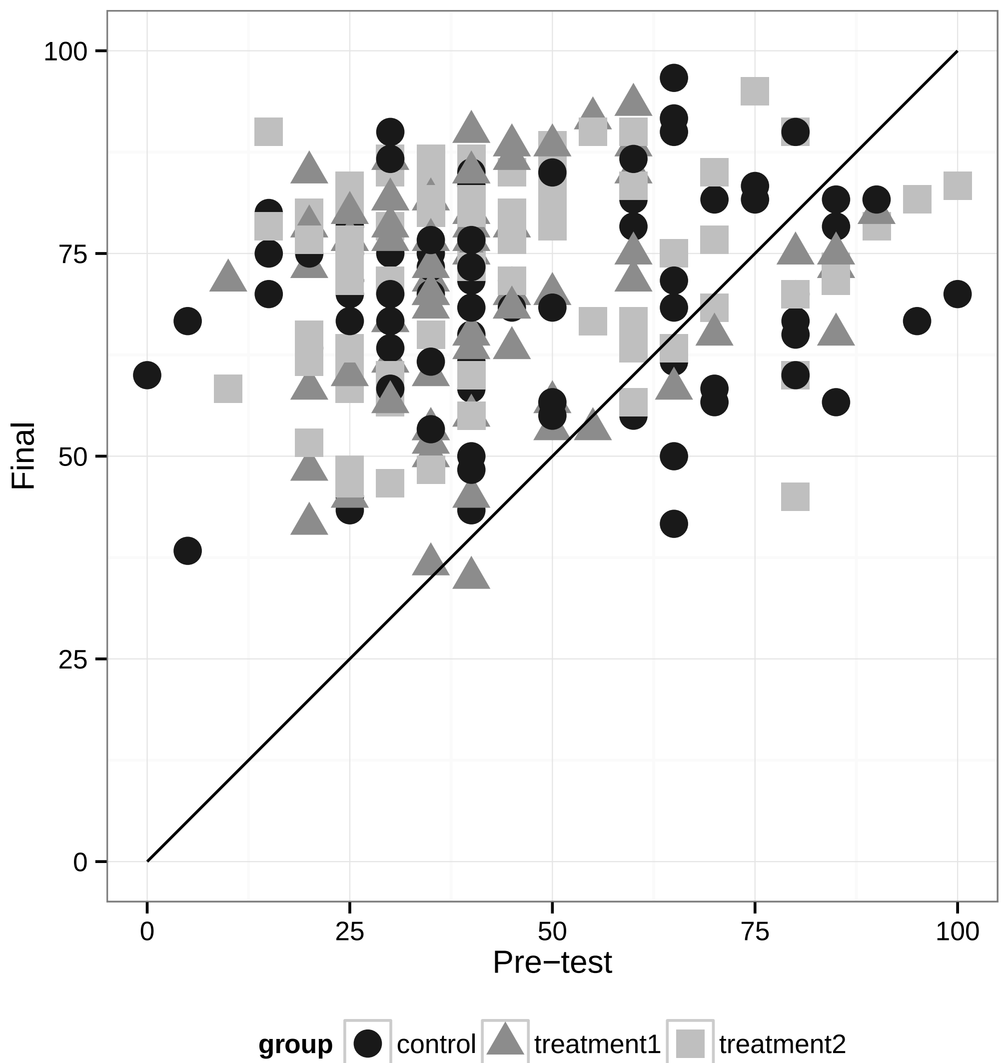


Figure 4.7: Midterm to pre-test improvement distribution

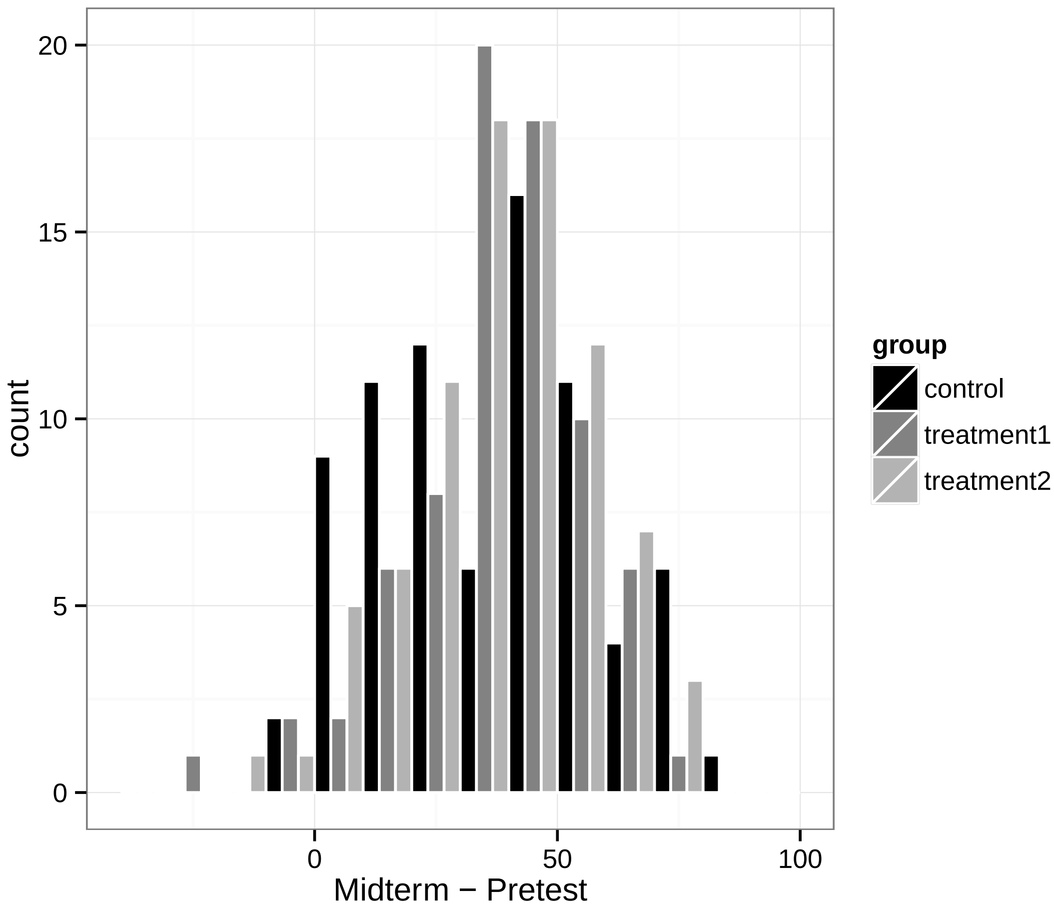


Table 4.1: Survey Answers by Treatment Group

	Control	Treatment 1	Treatment 2
Use of optional (ungraded) materials			
none	6	3	2
very little	14	12	13
some	20	25	35
most	15	7	10
all	2	1	3
Expected grade			
A	39	31	39
B+	12	13	20
B	4	4	4
B-	1	0	0
C	1	0	0
What did you think about the email?			
don't read email	1	0	1
don't remember receiving	45	10	11
didn't read	6	6	9
a scam	1	8	8
legitimate	4	24	34
Was the email clear?			
No	4	5	8
No Answer	5	0	0
Not sure	45	18	25
Yes	3	25	30
How long do you think it would take you to sign up?			
Not sure	34	11	17
less than 2 minutes	7	10	15
less than 10 minutes	15	21	26
more than 10 minutes	1	6	5
How long do you think it would take you to complete the Coursera course?			
No answer	0	0	0
Not sure	40	23	30
Less than 2 hours a week	0	0	0
Between 2 and 4 hours a week	0	0	0
More than 4 hours a week	0	0	0
Are you enrolled in the Coursera course?			
No	53	35	48
Yes	4	13	15
The reason you did not enroll is			
did not know about the existence of the course	49	10	13
email was confusing	3	10	22
not worth your time	5	15	21
Other:	5	15	8
No Answer	0	5	8
Minimum payment to try Coursera?			
$0	10	1	4
$5	4	4	1
$10	3	12	4
$25	18	7	15
$50	11	11	25
$75	5	4	2
$100	3	4	8
More than $100	1	3	3

Table 4.2: Regression Results, Coursera Enrollment and Students' Improvement

	Dependent variable			
	Improvement Midterm[a]	Enrolled in Coursera[b]	Improvement Midterm[a]	Improvement Final[c]
	OLS	*OLS*	*instrumental variable*	*instrumental variable*
	(1)	(2)	(3)	(4)
Enrolled in Coursera	6.615*		14.254	49.486*
	(3.710)		(21.256)	(29.586)
$10 for enrolling		0.124**		
		(0.055)		
$50 for 80% in Coursera		0.132**		
		(0.054)		
Constant	35.914***	0.051	34.869***	19.436***
	(1.372)	(0.039)	(3.179)	(4.424)
Observations	234	234	234	234
Residual Std. Error	19.498 (df = 232)	0.340 (df = 231)	19.675 (df = 232)	27.385 (df = 232)
F Statistic	3.180* (df = 1; 232)	3.697** (df = 2; 231)		

Notes: *p<0.1; **p<0.05; ***p<0.01
Sample: Students enrolled in PHYS 1060 in spring 2014 at the University of Virginia
[a] Difference between midterm and pre-test grades
[b] Dummy equal to 1 if the student is enrolled in Coursera
[c] Difference between final exam and pre-test grades

Table 4.3: Regression Results, Coursera enrollment and students' improvement. Restricted sample

	Dependent variable			
	Improvement Midterm[a]	Enrolled in Coursera[b]	Improvement Midterm[a]	Improvement Final[c]
	OLS	*OLS*	*instrumental variable*	*instrumental variable*
	(1)	(2)	(3)	(4)
Enrolled in Coursera	3.545		37.050*	60.140*
	(3.835)		(21.960)	(30.920)
\$10 for enrolling		0.171***		
		(0.064)		
\$50 for 80% in Coursera		0.115*		
		(0.064)		
Constant	28.790***	0.036	24.420***	10.010**
	(1.385)	(0.045)	(3.267)	(4.600)
Observations	161	161	161	161
Residual Std. Error	16.390 (df = 159)	0.332 (df = 158)	19.940 (df = 159)	28.070 (df = 159)
F Statistic	0.855 (df = 1; 159)	3.734** (df = 2; 158)		

Notes: *p<0.1; **p<0.05; ***p<0.01
Sample: Students enrolled in PHYS 1060 in spring 2014 at the University of Virginia that were in the pre-test top 75 percentile
[a] Difference between midterm and pre-test grades
[b] Dummy equal to 1 if the student is enrolled in Coursera
[c] Difference between final exam and pre-test grades

Bibliography

Akerlof, George A. 1991. "Procrastination and obedience." *American Economic Review* 81 (2):1–19.

Akinsola, Mojeed Kolawole, Adedeji Tella, and Adeyinka Tella. 2007. "Correlates of academic procrastination and mathematics achievement of university undergraduate students." *Eurasia Journal of Mathematics, Science & Technology Education* 3 (4):363–370.

Almasi, George S and Allan Gottlieb. 1989. *Highly parallel computing.* Benjamin-Cummings Publishing Co., Inc.

Anderson, Gary M and Walter Block. 1995. "Procrastination, Obedince, and Public Policy: The Irrelevance of Salience." *American Journal of Economics and Sociology* 54 (2):201–215.

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos. 2009. "Incentives and services for college achievement: Evidence from a randomized trial." *American Economic Journal: Applied Economics* :136–163.

Angrist, Joshua, Philip Oreopoulos, and Tyler Williams. 2010. "When opportunity knocks, who answers? New evidence on college achievement awards." Tech. rep., National Bureau of Economic Research.

Angrist, Joshua D and Jörn-Steffen Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

Ariely, Dan and Klaus Wertenbroch. 2002. "Procrastination, deadlines, and performance: Self-control by precommitment." *Psychological Science* 13 (3):219–224.

Bailey, Martha J and Susan M Dynarski. 2011. "Gains and gaps: Changing inequality in US college entry and completion." Tech. rep., National Bureau of Economic Research.

Banerjee, Abhijit V and Esther Duflo. 2014. "(Dis) organization and Success in an Economics MOOC." *The American Economic Review* 104 (5):514–518.

Barankay, Iwan. 2012. "Rank incentives: Evidence from a randomized workplace experiment." *Discussion papers* .

Barrow, Lisa, Lashawn Richburg-Hayes, Cecilia Elena Rouse, and Thomas Brock. 2012. "Paying for performance: The education impacts of a community college scholarship program for low-income adults." Tech. rep., Working Paper, Federal Reserve Bank of Chicago.

Barrow, Lisa and Cecilia E Rouse. 2013. "Financial incentives and educational investment: the impact of performance-based scholarships on student time use." Tech. rep., National Bureau of Economic Research.

Becker, Gary S. 1967. "Human Capital and the Personal Distribution of Income: An Analytical Approach (University of Michigan, Ann Arbor, Michigan)."

B.L. Beck, D.L. Milgrim, S.R. Koons. 2000. "Correlates and consequences of behavioral procrastination: The effects of academic procrastination, self-consciousness, self-esteem and self-handicapping." *Journal of Social Behavior and Personality* 15 (5):3–13.

Bong, Mimi and Richard E Clark. 1999. "Comparison between self-concept and self-efficacy in academic motivation research." *Educational psychologist* 34 (3):139–153.

Bowen, William G. 2013. "Academia Online: Musings (Some Unconventional)." .

Breslow, LB, David E Pritchard, Jennifer DeBoer, Glenda S Stump, Andrew D Ho, and DT Seaton. 2013. "Studying learning in the worldwide classroom: Research into edX's first MOOC." *Research & Practice in Assessment* 8:13–25.

Cacioppo, John T and Richard E Petty. 1982. "The need for cognition." *Journal of personality and social psychology* 42 (1):116.

Carbonaro, William. 2004. "Parent, Peer, and Teacher Influences on Student Effort and Academic Outcomes." In *ASA National Conference. San Francisco.*

———. 2005. "Tracking, students' effort, and academic achievement." *Sociology of Education* 78 (1):27–49.

Cheng, Cho Kin, Dwayne E Paré, Lisa-Marie Collimore, and Steve Joordens. 2011. "Assessing the effectiveness of a voluntary online discussion forum on improving students' course performance." *Computers & Education* 56 (1):253–261.

Choi, James J, David Laibson, Brigitte C Madrian, and Andrew Metrick. 2003. "Optimal defaults." *American Economic Review* :180–185.

Christensen, Gayle, Andrew Steinmetz, Brandon Alcorn, Amy Bennett, Deirdre Woods, and Ezekiel J. Emanuel. 2013. "The MOOC Phenomenon: Who Takes

Massive Open Online Courses and Why?" URL http://ssrn.com/abstract=2350964. Submitted for publication.

Chun Chu, Angela Hsin and Jin Nam Choi. 2005. "Rethinking procrastination: Positive effects of" active" procrastination behavior on attitudes and performance." *The Journal of Social Psychology* 145 (3):245–264.

Clark, Damon and Paco Martorell. 2014. "The signaling value of a high school diploma." *Journal of Political Economy* 122 (2):282–318.

Cohen, Arthur R, Ezra Stotland, and Donald M Wolfe. 1955. "An experimental investigation of need for cognition." *The Journal of Abnormal and Social Psychology* 51 (2):291.

Connolly, Marie. 2008. "Here comes the rain again: Weather and the intertemporal substitution of leisure." *Journal of Labor Economics* 26 (1):73–100.

Crotty, James Marshall. 2012. "Global Private Tutoring Market Will Surpass $102.8 Billion By 2018." http://tinyurl.com/ccduoj9. Accessed: 2014-06-03.

Einav, Liran and Jonathan D Levin. 2013. "The Data Revolution and Economic Analysis." Tech. rep., National Bureau of Economic Research.

Ellis, Albert and William J Knaus. 1977. *Overcoming procrastination: Or how to think and act rationally in spite of life's inevitable hassles*. Institute for Rational Living.

Fritzsche, Barbara A, Beth Rapp Young, and Kara C Hickson. 2003. "Individual differences in academic procrastination tendency and writing success." *Personality and Individual Differences* 35 (7):1549–1557.

Fryer, Roland G. 2011. "Financial incentives and student achievement: Evidence from randomized trials." *The Quarterly Journal of Economics* 126 (4):1755–1798.

———. 2013. "Information and Student Achievement: Evidence from a Cellular Phone Experiment." Tech. rep., National Bureau of Economic Research.

Garrison, D Randy, Terry Anderson, and Walter Archer. 2001. "Critical thinking, cognitive presence, and computer conferencing in distance education." *American Journal of distance education* 15 (1):7–23.

———. 2010. "The first decade of the community of inquiry framework: A retrospective." *The Internet and Higher Education* 13 (1):5–9.

Gerber, Alan S, Donald P Green, and Christopher W Larimer. 2008. "Social pressure and voter turnout: Evidence from a large-scale field experiment." *American Political Science Review* 102 (01):33–48.

Howell, Andrew J, David C Watson, Russell A Powell, and Karen Buro. 2006. "Academic procrastination: The pattern and correlates of behavioural postponement." *Personality and Individual Differences* 40 (8):1519–1530.

Hoxby, Caroline M. 2014. "The Economics of Online Postsecondary Education: MOOCs, Nonselective Education, and Highly Selective Education." Tech. rep., National Bureau of Economic Research.

Hungerford, Thomas and Gary Solon. 1987. "Sheepskin effects in the returns to education." *The review of economics and statistics* :175–177.

Jackson, Clement Kirabo. 2010. "A little now for a lot later: A look at a Texas advanced placement incentive program." *The Journal of Human Resources* 45 (3).

Johnson, Monica Kirkpatrick, Robert Crosnoe, and Glen H Elder Jr. 2001. "Students' attachment and academic engagement: The role of race and ethnicity." *Sociology of Education* :318–340.

Kizilcec, René F, Chris Piech, and Emily Schneider. 2013. "Deconstructing disengagement: analyzing learner subpopulations in massive open online courses." In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*. ACM, 170–179.

Knaus, William J. 2001. "Procrastination, blame, and change." *Journal of Social Behavior and Personality* 15 (5):153–166.

Lay, Clarry H. 1986. "At last, my research article on procrastination." *Journal of Research in Personality* 20 (4):474–495.

Levitt, Steven D, John A List, Susanne Neckermann, and Sally Sadoff. 2012. "The behavioralist goes to school: Leveraging behavioral economics to improve educational performance." Tech. rep., National Bureau of Economic Research.

Liyanagunawardena, Tharindu Rekha, Andrew Alexandar Adams, and Shirley Ann Williams. 2013. "MOOCs: A Systematic Study of the Published Literature 2008-2012." *International Review of Research in Open & Distance Learning* 14 (3).

Lloyd, Kenneth E and Norma J Knutzen. 1969. "A self-paced programmed undergraduate course in the experimental analysis of behavior." *Journal of Applied Behavior Analysis* 2 (2):125–133.

Marks, Helen M. 2000. "Student engagement in instructional activity: Patterns in the elementary, middle, and high school years." *American educational research journal* 37 (1):153–184.

Martinez, Ignacio. 2014a. "The effects of informational nudges on students' effort and performance: Lessons from a MOOC."

———. 2014b. "Never put off 'till tomorrow?"

Martinez, Ignacio and Paul Diver. 2014. "MOOCs as a Massive Research Laboratory: Opportunities and Challenges." .

Mas, Alexandre and Enrico Moretti. 2009. "Peers at Work." *American Economic Review* 99 (1):112–145.

McCarney, Rob, James Warner, Steve Iliffe, Robbert van Haselen, Mark Griffin, and Peter Fisher. 2007. "The Hawthorne Effect: a randomised, controlled trial." *BMC medical research methodology* 7 (1):30.

Michinov, Nicolas, Sophie Brunot, Olivier Le Bohec, Jacques Juhel, and Marine Delaval. 2011. "Procrastination, participation, and performance in online learning environments." *Computers & Education* 56 (1):243–252.

Newmann, Fred M. 1992. *Student Engagement and Achievement in American Secondary Schools*. ERIC.

O'Donoghue, Ted and Matthew Rabin. 2001. "Choice and procrastination." *The Quarterly Journal of Economics* 116 (1):121–160.

Owens, Anthony M and Ian Newbegin. 1997. "Procrastination in high school achievement: A causal structural model." *Journal of social Behavior & personality* .

Papay, John P, Richard J Murnane, and John B Willett. 2011. "How performance information affects human-capital investment decisions: the impact of test-score labels on educational outcomes." Tech. rep., National Bureau of Economic Research.

Rabin, Matthew. 1998. "Psychology and economics." *Journal of economic literature* 36 (1):11–46.

Rotenstein, Aliza, Harry Z Davis, and Lawrence Tatum. 2009. "Early Birds versus Just-in-Timers: The effect of procrastination on academic performance of accounting students." *Journal of Accounting Education* 27 (4):223–232.

Rothblum, Esther D, Laura J Solomon, and Janice Murakami. 1986. "Affective, cognitive, and behavioral differences between high and low procrastinators." *Journal of counseling psychology* 33 (4):387.

Semb, G, DM Glick, and RE Spencer. 1977. "Programmatic research: An analysis of student withdrawals in behavioral systems of instruction." In *Fourth national conference on behavior research and technology in higher education, Pittsburg.*

Siegfried, John J. 2001. "Principles for a successful undergraduate economics honors program." *The Journal of Economic Education* 32 (2):169–177.

Smith, Jonathan. 2013. "Peers, Pressure, and Performance at the National Spelling Bee." *Journal of Human Resources* 48 (2):265–285.

Solomon, Laura J and Esther D Rothblum. 1984. "Academic procrastination: Frequency and cognitive-behavioral correlates." *Journal of Counseling Psychology* 31 (4):503.

Spence, Michael. 1973. "Job market signaling." *The quarterly journal of Economics* 87 (3):355–374.

Sullivan, Paul. 2010. "As Private Tutoring Booms, Parents Look at the Returns." `http://tinyurl.com/mmccafl`. Accessed: 2014-06-03.

Sunstein, Cass R and Richard Thaler. 2008. *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press New Haven.

Thaler, Richard H and Cass R Sunstein. 2008. *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Trost, Steve and Djavad Salehi-Isfahani. 2012. "The effect of homework on exam performance: Experimental results from principles of economics." *Southern Economic Journal* 79 (1):224–242.

Tuckman, Bruce W. 1998. "Using tests as an incentive to motivate procrastinators to study." *The Journal of experimental education* 66 (2):141–147.

Tversky, Amos and Daniel Kahneman. 1985. "The framing of decisions and the psychology of choice." In *Behavioral decision making*. Springer, 25–41.

Van Eerde, Wendelien. 2003. "A meta-analytically derived nomological network of procrastination." *Personality and Individual Differences* 35 (6):1401–1418.

Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28 (2).

Vytlacil, Edward and Nese Yildiz. 2007. "Dummy endogenous variables in weakly separable models." *Econometrica* 75 (3):757–779.

Wooldridge, Jeffrey. 2012. *Introductory econometrics: A modern approach*. Cengage Learning.

Zhang, Yong and Richard Buda. 1999. "Moderating effects of need for cognition on responses to positively versus negatively framed advertising messages." *Journal of Advertising* 28 (2):1–15.