Using learning analytics to evaluate design changes in MOOCs : A case study on assessing course pacing

by

Ahmed Bilal

MBA, Massachusetts Institute of Technology (2018)

Submitted to the System Design and Management Program
in partial fulfillment of the requirements for the degree of

Master of Science in Engineering and Management

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2019

Signature redacted

Author ....

System Design and Management Program
August 23, 2019

Signature redacted

Certified by.. ....

Chris Caplice
Executive Director, MIT Center for Transportation & Logistics
Thesis Supervisor

Signature redacted

Certified by.. ....

Eva Ponce
Executive ~~Director, MITx Micro~~Masters in Supply Chain Management
Thesis Supervisor

Signature redacted

Accepted by ....

Joan Rubin
Executive Director, System Design and Management Program

Using learning analytics to evaluate design changes in MOOCs : A case study on assessing course pacing

by

Ahmed Bilal

Submitted to the System Design and Management Program
on August 23, 2019, in partial fulfillment of the
requirements for the degree of
Master of Science in Engineering and Management

Abstract

Experimentation on course design in MOOCs can determine causal factors that promote learning and can identify aspects of the course where revision is needed. The presence of heterogeneous samples of learners, the difficulty of defining success metrics, and the lack of shared cross-course data are few of the challenges course designers face to evaluate MOOCs.

In this thesis, we present a data-driven framework to evaluate design changes in MOOCs. We explore a change from multiple angles -process, proficiency, and perception- and apply various analytical methods -temporal, causal and predictive- to map out the outcome of instruction along multiple dimensions of learning.

We demonstrate the application of this framework by evaluating course pacing on a repeated run of a supply chain MOOC by MITx. Self-pacing caused completion rate (-6%), pass rate (-10%), and engagement score (-7%) to drop, although students' satisfaction with course remained unchanged. The impact of pacing on students' outcome was not uniform with some experiencing no change while others encountering a steep fall. The most striking difference was seen in the longitudinal trajectories, with instructor-paced students mostly taking the same trajectory and self-paced students pursuing their own individually paced trajectories. We showed that these trajectories are correlated with student grade, and students with certain characteristics are inclined to pursue a specific trajectory.

From these and other observations, we were able to provide practical guidance to course designers on what instructional materials and practices are satisfactory and where change is needed.

Thesis Supervisor: Chris Caplice
Title: Executive Director, MIT Center for Transportation & Logistics

Thesis Supervisor: Eva Ponce
Title: Executive Director, MITx MicroMasters in Supply Chain Management

Acknowledgments

This body of work has benefited from the support and encouragement of many people. Without their help, I would not have been able to craft the following work. I am genuinely indebted to them all. I would like to mention, particularly:

- Chris and Eva, my advisors, for inspiring me with their passion for online education and for giving me the opportunity to work on this project. Together, they are changing the landscape of online education, and I can't thank them enough for allowing me to experience this transformation first hand. I will also like to thank them for giving me the freedom as well as the resources to explore possible research directions. Their guidance and mentorship allowed me to become a better student and professional throughout this process. Without their help, none of this would be possible.

- Alexis and Inma, for their insightful suggestions and constructive feedback. Alexis significantly improved the way I framed my research. Inma's intuition for the courses drove my research in the right direction.

- Connor, David, Sergio, Andrea, and Arthur for their support and guidance. It has been a pleasure to work with them.

- The MBA, SDM and SCM faculty, staff, and cohort, for offering so much inspiration and knowledge during my time here.

- Nehal, my wife, for her consistent love and support throughout the past year.

- Taimur, my son, for reminding me there is a life outside of MIT.

- Most importantly, my parents, for their unconditional love and never-ending support. I am here today because of them.

Contents

1 Introduction **17**

1.1 What is course design and why is it important? 17

1.2 Course design in MOOCs 18

1.3 Evaluating course design in MOOCs 20

1.4 Challenges of evaluating course design in MOOCs 20

1.5 Research Questions 23

1.6 Contributions 24

1.7 Thesis Overview 25

2 Literature Review **27**

2.1 On Experimental Research in MOOCs 27

2.1.1 Plug-in experiments 27

2.1.2 Complex design experiments 28

2.2 On Course Pacing in MOOCs 29

2.2.1 Synchronous course delivery 29

2.2.2 Asynchronous course delivery 30

2.2.3 Semi-synchronous course delivery 30

2.2.4 Related Work on Pacing 31

3 Framework to evaluate course design in MOOCs **33**

3.1 Process Measures 34

3.2 Proficiency Measures 35

3.3 Perception Measures 35

3.4 Analytical Methods Employed 36
3.4.1 Data Acquisition and Integration 36
3.4.2 Identifying sub-populations 37
3.4.3 Static and Dynamic Analysis 37
3.4.4 Causal and Predictive Methods 37
3.5 Defining Student success 37
3.5.1 Completion 38
3.5.2 Pass rate 38
3.5.3 Engagement 38
3.6 Data-driven Framework to evaluate course design changes 40

4 Case Study: Evaluating pacing in MOOCs 43

5 Data Collection and Feature Engineering 51
5.1 Course Structure and Content 51
5.1.1 Course Structure 51
5.1.2 Course Contents 52
5.1.3 Types of Learners 53
5.2 Data Systems and Sources 54
5.2.1 Learner Demographics Data 54
5.2.2 Survey Data 54
5.2.3 Activity Data 58
5.2.4 Course metadata 58
5.2.5 Achievement data 58
5.3 Summary Statistics 59

6 Process Measurement I: Static analysis of events taking place in MOOCs 67
6.1 Course Activity and Engagement 68
6.2 Grade and Engagement 69
6.3 Course Content-Type and Engagement 71

6.4 Performance and engagement 74
6.5 Dropout and Engagement 79

7 Process Measurement II: Temporal analysis of events taking place in MOOCs 83

7.1 Exploratory Temporal Analysis 83
7.2 Clustering Method 85
7.3 Engagement Trajectory for Completers 90
7.3.1 Completers in Instructor-paced course 90
7.3.2 Completers in self-paced course 91
7.4 Engagement Trajectory for Dropouts 95
7.4.1 Dropouts in Instructor-paced course 95
7.4.2 Dropouts in self-paced course 95
7.5 Clusters and Demographics 96
7.5.1 Method 96
7.5.2 Result 99
7.6 Clusters and Outcome 100
7.6.1 Method 100
7.6.2 Result 101
7.7 Conclusion 101

8 Proficiency Measurement: Impact of course pacing on student success 105

8.1 Method 106
8.2 Result 108
8.3 Heterogeneous treatment effects 110
8.4 Conclusion 112

9 Perception Measurement: Impact of course pacing on students' perception 113

9.1 Interest in the subject 113

9.2 Interpersonal outcomes 114
9.3 Intrapersonal outcomes 115
9.4 Course Content 115
9.5 Certification 117
9.6 Conclusion 117

10 Impact of course pacing on predictive models 119

10.1 Problem Formulation 119
10.2 Method 120
10.3 Transferability of predictive model 121
10.4 Factors correlated with dropouts 122
10.5 Conclusion 125

11 Conclusion, Recommendation and Future Research 127

11.1 Conclusion 127
11.1.1 Framework to evaluate design changes in MOOCs 127
11.1.2 Case Study: Evaluating pacing in MOOCs 128
11.2 Recommendation 129
11.2.1 When to use Instructor-pacing? 129
11.2.2 When to use student-pacing? 130
11.2.3 How to improve success in student-pacing? 130
11.3 Limitations and Future Research 133

A Tables 135

List of Figures

4-1 History of Supply Chain Analytics (SC0x) MOOC from 2016-present 45

6-1 Median count of actions of students with a given final grade in instructor-paced and self-paced version of SC0x. 72
6-2 Probability distribution of Core Lectures and Recitation lectures for completers in the instructor-paced and self-paced version of SC0x . . 75
6-4 Probability distribution of Extra PP and Exam Prep for completers in the instructor-paced and self-paced version of SC0x 76
6-5 Engagement trajectory of problem activities separated by problem type for completers in self-paced version of SC0x 76
6-6 Engagement trajectory of video activities separated by chapter/module type for completers in instructor-paced and self-paced version of SC0x 77
6-7 Probability distribution of Video, Assessment and Forum activities of students with course grade of A and F in instructor-paced and self-paced version of SC0x 80
6-8 Probability distribution of percentage of videos consumed by completers and dropouts across the versions of SC0x 81
6-9 Probability distribution of percentage of ungraded problems submitted by completers and dropouts across the versions of SC0x 81

7-1 Weekly distribution of number of unique videos watched for completers and dropouts in instructor-paced and self-paced version of SC0x . . . 86
7-2 Weekly distribution of number of problems submitted for completers and dropouts in instructor-paced and self-paced version of SC0x . . . 87

7-3 Engagement trajectory of completers in instructor-paced and self-paced versions of SC0x 88
7-4 Engagement trajectory of dropouts in instructor-paced and self-paced versions of SC0x 89
7-5 Summary of clustering method used to determine longitudinal engagement trajectories in SC0x 90
7-6 Clustering engagement trajectories for completers in instructor-paced versions of SC0x 93
7-7 Clustering engagement trajectories for completers in self-paced version of SC0x 94
7-8 Clustering engagement trajectories for dropouts in instructor-paced versions of SC0x 97
7-9 Clustering engagement trajectories for dropouts in self-paced version of SC0x 98

10-1 Comparison of Test AUC scores of various classifiers 122
10-2 Test and Train Score of the predictive models over weeks 123
10-3 Visualization of SHAP values for models trained on data from instructor-paced course and self-paced course 126

List of Tables

3.1 Data-driven Framework to evaluate design changes in MOOCs 41

4.1 Comparison of course content across the versions of SC0x 45
4.2 Comparison of demographic features across the versions of SC0x . . . 47
4.3 Comparison of enrollment and verification timing across the versions of SC0x 48
4.4 Comparison of entrance survey features across the versions of SC0x . 49

5.1 List of demographic features extracted from the platform 55
5.2 List of features extracted from entrance survey 56
5.3 List of features extracted from exit survey 57
5.4 List of general course features extracted from the platform 59
5.5 List of video features extracted from the platform 59
5.6 List of forum features extracted from the platform 60
5.7 List of problem features extracted from the platform 60
5.8 List of video metadata extracted from the platform 61
5.9 List of problem metadata extracted from the platform 61
5.10 List of achievement data extracted from the platform 61
5.11 Summary Statistics of the demographic features for SC0x 64
5.12 Summary statistics of clickstream data for SC0x 65

6.1 Comparison of video activities across the versions of SC0x 69
6.2 Comparison of assessment activities across the versions of SC0x . . . 70
6.3 Comparison of forum activities across the versions of SC0x 70

7.1 Output of logistic regression models predicting cluster type using demographic features for completers in self-paced version of SC0x . . . 102
7.2 Output of logistic regression models predicting cluster type using demographic features for completers in self-paced version of SC0x . . . 103
7.3 Output of logistic regression models predicting course score using trajectory types for completers in self-paced version of SC0x 104

8.1 Summary of elements used in potential outcomes framework 107
8.2 Student success metrics for all runs of SC0x 109
8.3 Impact of self-pacing on student success metrics using all previous instructor-paced versions (2017-present) 109
8.4 Impact of self-pacing on student success metrics using instructor-paced versions starting from 2018 (2018-present) 109

9.1 Students' perception of interest in the subject 114
9.2 Students' perception of interpersonal gains from the course 115
9.3 Students' perception of intrapersonal gains from the course 116
9.4 Students' perception of quality of the course 116
9.5 Students' reasons for pursuing verified certificate 117

10.1 Taxonomy of predictive model used in our stud 121
10.2 Comparison of Test AUC scores of various classifiers 121

A.1 Clusters and their demographic profile for completion metrics 136
A.2 Clusters and their demographic profile for pass rate metrics 137
A.3 learners' gender ratio in all runs of SC0x 138
A.4 learners' age ratio in all runs of SC0x 138
A.5 Country of residence of learners in all runs of SC0x 138
A.6 Level of education of learners in all runs of SC0x 139
A.7 HDI of country of residence of learners in all runs of SC0x 139
A.8 Englisg language index of learners in all runs of SC0x 139
A.9 Enrollment registration of learners in all runs of SC0x 139

A.10 Verification registration of learners in all runs of SC0x 140
A.11 EdX joining data of learners in all runs of SC0x 140
A.12 Employment status of learners in all runs of SC0x 140
A.13 Language proficiency of learners in all runs of SC0x 140
A.14 Familiarity with topic of learners in all runs of SC0x 141

Chapter 1

Introduction

"There is a new world unfolding, and everyone will have to adapt."

-L. Rafael Reif, M.I.T. president, on MOOC in 2013

1.1 What is course design and why is it important?

Course design refers to content, duration, sequencing, and other relevant design characteristics of the elements of a course (including course requirements, possible pathways, nature of the content, expected learning outcomes, assessment structure, certifications, etc.) [55]. Broadly speaking, course design can include all activities or processes that an instructor carries out to plan for successful student outcome [1].

An effective course design has shown to increase students' engagement, learning, and satisfaction with the course [24]. Carolyn Fellahi redesigned her residential psychology course from a traditional lecture-driven method to a one that includes multiple modes of learning and found an increase in the score for five of the six learning outcomes [22]. Bill Weeks overhauled his on-campus programming theory course from a lecture-assessment driven method to a one that uses team-based learning and found drastic improvements in students' morale and satisfaction [72]. Black and others measured the impact of course design on students' outcomes in a residential marketing course and found course design to influence students' perception and performance [6]. Kim and others showed that a well-organized course with clear goals, relevant assignments, and actively facilitated discussions increased cognitive and critical thinking abilities of students [29]. Thomas Moore argued that a bad course design could demoralize students while good design can have an opposite effect, going so far as to claim that "good course design trump almost everything else" [51].

In spite of the importance of course design, creating efficient course design is not straightforward and demands a lot of extensive planning and reflection. Researchers have proposed various frameworks to help instructors with the design process. Universal design for learning provides a framework to help instructors develop a course that is accessible to everyone [62]. Integrated course design presents a formal structure to help instructors determine course goals, activities, and assessments when creating an integrated course design [24]. Backward course design begins by identifying the desired outcomes of a course and then planning teaching activities to achieve those outcomes [15].

Our aim in this thesis is not to comment on how to design an effective course or what design framework to use; instead, we are interested in developing strategies to help instructors evaluate course design. Irrespective of what change is made or what method is used, one needs to perform course evaluation to measure the impact of a design change on the students' outcomes and the course. Observations from the evaluation process can identify aspects of the course where revision is desirable, providing guidance and inspiration for instructors to kick in the design process again. A course design is, therefore, a cyclic process in which the breadth and quality of the evaluation process connect the cycles together [1].

1.2 Course design in MOOCs

Massive open online courses (MOOCs) have been advocated as a solution to lift people out of poverty, to unlock a billion more brains and to disrupt higher education [26]. At the same time, MOOCs have been criticized with researchers calling out their low retention rate, declining enrollment, and general affluence of students as reasons of why MOOCs don't work [58]. Regardless of which side one supports, no one can deny the enthusiasm and the participation that MOOCs have generated over the last six years. Even if MOOCs don't transform the higher education as was perceived initially, they are here to stay and will continue to serve millions of learners around the globe [42].

The digital and open nature of the MOOCs makes course design even more important. Instructors can not respond to students individually and have to embed architectural changes in the course to keep their diverse audience engaged [26]. The digitization of courses adds additional complexity in the design process. Instructors need to choose how do students interact with the course and how is the material presented.

Experimentation on course design started soon after MOOCs became popular. The experimental research on course design to date has centered around two broader themes: "plug-in experiments" and "complex design experiments" [57]. "Plug-in experiments" relates to experiments where interventions are made within a course to boost motivation, memorization, or other common facets of learning. "Complex design experiments" involves experiments that evaluate the overarching pedagogical process in MOOCs. These experiments examine theory-based changes to course content and course structure. In this thesis, we focus mostly on complex design experiments, although the strategies discussed are transferable to plug-in experiments as well.

Finally, we argue that the course design offers a practical way forward to improve the high dropout rate of MOOCs. Despite several years of investment in this topic, the completion rate of MOOCs did not increase [58]. Most research on improving dropout has centered on demonstrating a positive correlation between student activity and success [66, 69, 53]. However, these studies have failed to provide instructions on what can be done to improve learning [57]. Experimentation on course design can determine causal factors that promote learning and can identify aspects of the course where revision is desirable. Given that such changes are in the hands of instructors, design changes can improve teaching and reduce dropout.

1.3 Evaluating course design in MOOCs

MOOCs offer an unprecedented opportunity for educationalists and researchers to evaluate course design at a scale and depth never done before. It is for this reason that the chief executive of EdX has dubbed MOOCs as the "particle accelerator for learning" [68]. The potential of MOOCs as a research instrument can be attributed to its two novel features: first being its ability to collect big, diverse data and the second being its capacity to perform field experiments [33]. On the data front, researchers can now collect learner data that is extensive and comprehensive. This data comes from thousands of learners that have diverse demographics and socio-economic background, making it possible for educators to test scientific theories on a broader population. On the experimentation front, researchers can now conduct field experiments rapidly and economically. These experiments could be randomized or natural and can be evaluated quickly, helping researchers to test causal relationships between course interventions and student outcomes.

Despite these affordances, most studies on course design in MOOCs did not produce significant results [33]. The presence of heterogeneous samples of learners, the difficulty of defining success metrics, and the lack of shared cross-course data are few of the challenges researchers have faced in evaluating MOOCs.

In this thesis, we present a framework to evaluate design changes in MOOCs and demonstrate the application of this framework by evaluating pacing structure on a repeated run of the same course.

1.4 Challenges of evaluating course design in MOOCs

The impact of course design on learners' outcomes is little understood. In this section, we explore the challenges and the limitation faced by educators to carry this stream of work:

1. Barriers to access shared MOOCs data

One limitation to evaluate course design in MOOCs has been the lack of shared course data. A researcher can conduct thorough investigations within a course but needs data from multiple courses to make a robust claim of the impact of particular course design on student success [57]. There are political, regulatory, and technological barriers to share data. Therefore, most researchers have either explored design changes in a single run of the course or examine design changes across courses that differ in many dimensions. Both these methods limit one's ability to draw reasonable inference about the impact of course design in a MOOC.

2. Limitation to conduct randomized experiments

Another challenge is the limitation to conduct a randomized control trial to evaluate pedagogical approaches in MOOCs. The gold standard to determine the efficacy and effectiveness of interventions and policy programs have long been the randomized controlled trial (RCT). Researchers, from the earliest MOOCs, have implemented several such trials but the scope and the reach of these experiments have been narrow [13, 34, 40, 47, 4]. These experiments are limited and do not challenge the pedagogical process. However, researchers are increasingly asking structural questions for which the RCT may not be a practical (or ethical) option. Consider our study, where we are comparing the student success in a self-paced course with an instructor-paced course. There are limitations to conducting a full-scale RCT here: Students may not want to be randomized, or only atypical participants may be willing to be randomized. Such experiments also raise ethical issues and require additional resources to run two courses in parallel.

3. Limited sources of data

Most research on course improvement in MOOCs has relied on self-report measures from course surveys with relatively low response rates. Surveys can tell a lot about student's motivation and their perceptions about the change. However, they introduce self-selection into the study, which tends to skew the sample towards more committed learners who may respond to the treatment differently from those who opted against taking the survey. The other stream of research has focused exclusively on clickstream data. Clickstream activity captures student's behavioral activity in a course at a fine-grained level. However, they give very little understanding of the student's thoughts and intent.

4. Learners varying motivation to enroll in MOOCs

Another limitation is the presence of heterogeneous samples of learners in MOOCs. Koller [14] and Reich [56] showed that learners come to MOOCs with different intent and motivation. Recent field experiments on design changes in MOOCs had mostly produced inconclusive research with researchers attributing the lack of convincing results in these experiments to the interaction of interventions with the diverse motivations of individual students [60, 52, 18, 34] . Consider our study where we conditioned the experiment on learners that already had shown firm intention to complete the course by paying a course fee, but even among them, we observed varying motivation and intent. A change is likely to impact these learners in different ways, and one must take individual differences into account to determine the true effect of the change.

5. Lacking outcome metrics

Finally, researchers have struggled to determine what outcome to use to measure learning. Existing assessment structures such as completion and clicking are limited and do not support inferences about learning. Reich [57] and Savi [65] identified various issues with existing assessment statistics and stressed to focus on learning rather than mere clicking. We tend to agree as these basic statistics obscure the underlying goal of many changes and interventions. Consider a recent change in one of our course where we introduced an open response assignment (essay response and peer-graded component)and measured its impact on student success. It will be naive to expect to see an increase in completion from such a course change, but the hope is that these new type of assessments can increase conceptual understanding or expert thinking, helping learners to be successful in their professional or academic careers.

1.5 Research Questions

While there exist challenges to study course design in MOOCs, the benefits of evaluating design changes are many folds. An evaluation can measure how learners respond to change, telling us more about their motivation and intent. Importantly, an evaluation can identify aspects of the course where revision is desirable.

Considering the importance of evaluating course design in MOOCs and limited research in this space, the first research question that guided our agenda is as follows:

RQ1: How can we evaluate course design changes in MOOCs? What dimension needs to be considered, and how can we operationalize the process?

In this thesis, we present a data-driven framework to evaluate complex design changes in MOOCs. We explore a change from multiple angles -process, proficiency, and perception- and apply various analytical methods -temporal, causal and predictive- to map out the outcome of instruction along multiple dimensions of learning. Such an approach goes beyond the traditional viewpoint of comparing courses with a single monolithic metric (completion, certification, etc.) and provides instructors with practical guidance on what instructional materials and practices are satisfactory and where revision is required.

We demonstrate the application of this framework by evaluating course pacing on a repeated run of a MOOC by MITx. Pacing refers to how a cohort of learners engages with the material in the course.

Our analysis on pacing is based on an introductory course on supply chain (SC0x - Supply Chain Analytics) by MIT that has been offered six times since 2016. The first five-run of the SC0x were instructor-paced. These courses followed a set schedule where lessons were released weekly, and assignments were due every week. Students take a scheduled midterm and a final exam. In 2019, the pacing structure of SC0x was changed to self-paced. With this change, all of the course materials were released as soon as the course started. Learners were able to watch videos and practice problems through the course at their speed. While there is no midterm, the self-paced course did have a scheduled final exam. Apart from the pacing, the content and other aspects of the course structure did not change.

Given the pedagogical uniformity between the two versions, the self-paced version of SC0x provides an exciting opportunity to evaluate the impact of course pacing on students' performance.

To investigate the effect of student-pacing on student success, we explored the following research questions:

RQ2: How do students behave and act across the two versions of SCOx?
RQ3: Is there any impact of course pacing on students performance?
RQ4: What are the perceived effects of course pacing on students' learning experience?

Our analysis only considers verified learners. Verified learners are those that have paid a fee to demonstrate their intent to earn a credential with labor market value. We focus on verified learners because 1) these learners have access to all of the content in the course including exams, enabling us to perform a comprehensive review of the change and 2) Even among verified learners, the completion rate has remained low at around 50%, indicating that there are plenty of opportunities to support these learners.

1.6 Contributions

The thesis aims to provide a data-driven framework to help faculty and course designers evaluate course design changes in MOOCs. In an intent to remain thorough, we explore a change from multiple angles. Our contributions in this thesis include the following:

- Formally proposed a framework to evaluate course design changes in MOOCs and demonstrated the application of the framework by evaluating course pacing in MOOCs.
- Defined an engagement score that captures students' engagement in MOOCs across multiple dimensions.
- Integrated clickstream, survey, course meta, and achievement data to examine student's interaction during MOOCs.
- Developed a temporal classification method to identify longitudinal engagement trajectories in MOOCs.
- Applied causal inference methods to estimate the impact of course design on students' performance.
- Used machine learning methods to understand the impact of course design on dropouts.
- Shared guidelines for faculty and course designers involved with running or creating open online courses.

1.7 Thesis Overview

The remainder of this thesis is organized as follows:

- Chapter 2 reviews the literature relevant to our topic. We begin this chapter by reviewing existing work on experimental research in MOOCs, and later, we explore previous work on course pacing in MOOCs.
- Chapter 3 presents our methodology and outlines the basic components of our framework. We also define the engagement score in this chapter and identifies a possible way to operationalize it.

- Chapter 4 outlines the experimental setup that we use to explore pacing in MOOCs.

- Chapter 5 reviews our approach to data collection. We begin this chapter by presenting details on the course structure and content, and later, we share sources of data and the list of covariates extracted from the online platform.

- Chapter 6 and 7 examine events taking place in the course and compare them across both the versions of the course. In chapter 6, we explore events from a static point of view, i.e., we are concerned with the final state of the activity and not what path was taken to reach to that final state. In chapter 7, we present a temporal analysis of learner's engagement, i.e., we are interested in tracking learners' behaviors over time (weeks) and investigate if indeed a shift to the student-pacing has changed this behavior.

- Chapter 8 investigates the impact of course pacing on student success. We employ causal inference methods to draw a causal connection between pacing and student success metrics.

- Chapter 9 provides more explanation on the mechanism behind the dropout across both the versions. We employ predictive modeling methods to identify factors that are correlated with dropout. We also explore the transferability of predictive models between the two versions of the same course.

- Chapter 10 examines student perception of the design change.

- Chapter 11 summarizes the key findings of the thesis and provide guidelines for course developers. We also identify the limitations of our research and share possible future directions.

Chapter 2

Literature Review

In this chapter, we review the literature relevant to our topic. We begin this section by reviewing existing work on experimental research in MOOCs. Later, we zoom into one type of complex design change in a MOOC: course pacing.

2.1 On Experimental Research in MOOCs

The experimental research in MOOCs to date has centered around two broader themes: "plug-in experiments" and "complex design experiments" [57]. In this thesis, we focus on the later, although our methods apply to the former, as well.

2.1.1 Plug-in experiments

"Plug-in experiments" relates to experiments where interventions are made within a course to boost motivation, memorization, or other common facets of learning [57]. Anderson and others [3] investigated the impact of "virtual badges" on engagement in discussion forums. They employed a causal inference framework to control for the heterogeneous population within MOOCs. An increase in forum activities was observed for some students, but overall, the results were limited. On a similar line, Coetzee and others [41] experimented to explore the effect of follow-up emails on participation in discussion forums. Similar to Anderson, they also leveraged mathematical techniques from causal inference method to account for the heterogeneous population of MOOCs. They witnessed increased participation from already active students but found no impact on other students. Kizilcec and others [37] examined the impact of encouragement emails on engagement in discussion forums. They split learners into three groups and plotted forum posts across the number of weeks for each of the group. They found encouragements to be ineffective in motivating learners to participate. Renz and others [59] explored the impact of emails on lecture and forum participation in MOOCs. They introduced an A/B testing framework in MOOCs but found inconclusive results for the experiment. Borella and others [7] combined predictions with email interventions to motivate at-risk students to participate in exams. They used prior data of the course to predict students that are at the risk of dropout. Similar to Kizilcec, they found emails to be an ineffective tool to motivate students. Apart from emails, others have used surveys and peer discussion as an intervention to improve social engagement in the course [35, 10]. In summary, the focus of most research on "Plug-in experiments" has been to increase forum participation in MOOCs. Emails have been the most common interventional tools in these experiments with some researchers experimenting with incremental in-course designs. The most common methodology has been A/B test with some researchers combining it with causal inference framework to account for heterogeneous populations in MOOCs. Overall, the results of these experiments have yielded small or non-significant results.

2.1.2 Complex design experiments

"Complex design experiments" involves experiments that evaluate the overarching pedagogical process in MOOCs [57]. These experiments examine theory-based changes to course content and course structure. The literature on "Complex design experiments" is limited. Renz [60] introduced "onboarding videos" to familiarize learners with course structure and content but found no increase in engagement metrics. Kizilcec [34] and others examined the impact of instructor's face in a video on information retention and attention. They conducted their experiments outside MOOC on a relatively small group and used surveys to capture learners perception of the intervention.

No impact on engagement was observed, although learners did prefer watching videos with the instructor's face. Davis and others [18] investigated the effect of weekly reflection and course competition. They asked learners to summarize the content and next-week plan at the end of each week and measured engagement with and without this intervention. They observed no impact on completion and engagement. Fisher [25] studied the impact of curriculum on learning in a fascinating study on a MOOC course on copyright laws. He randomly assigned students to one of the two curricula—one based on U.S. case law and the other on global copyright issues—and use data from various source to evaluate the curricula. Overall, there have been limited studies on complex course design in MOOC with most studies founding no significant improvements in learning outcomes. The interventions in "Complex design experiments" are complex where design changes had to be baked in the architecture of MOOCs. The analytical methods and outcomes metrics vary in each case. The lack of standardization in evaluating these changes restricts the application of these experiments on other courses.

2.2 On Course Pacing in MOOCs

Pacing refers to how a cohort of learners engages with the material in the course [20].

2.2.1 Synchronous course delivery

A course is said to be synchronous if the learners have to follow a set schedule to go through the course material. That is, the course has a fixed start and end date; the videos need to be watched by the learners at the same time, and there are a series of due dates for assignment and exams along the way [52]. The benefit of synchronous design is that it fosters a sense of community by encouraging students to progress as a cohort. Most traditional courses are synchronous.

2.2.2 Asynchronous course delivery

On the flip side, a course is said to be asynchronous if the learners can engage with the course material at any pace. That is, there is no start and end date; the learners can watch videos at any pace, and there are no due dates of the assignments and exams [52]. The benefit of asynchronicity is that it offers learners the flexibility to participate in a course at any pace. Courses that are perfectly asynchronous includes MIT Open Course Ware, Khan Academy.

2.2.3 Semi-synchronous course delivery

MOOCs, on the contrary, are semi-synchronous. That is, they enable learners to go through a course at their own pace with certain restrictions. The intensity of restrictions varies a lot within MOOCs, with some courses prioritizing flexibility while others lean more towards community development. The trade-off between asynchronicity and synchronicity in a MOOC is a design choice, and it is up to the course designer to determine what decision to make. Nevertheless, the pacing is one of the distinctive features of MOOCs, and the level of synchronicity can fundamentally alter the structure of the course. Broadly speaking the pacing in MOOCs can be further divided into two sub-categories. Both of these categories are semi-synchronous, but the intensity of synchronicity varies between them.

Instructor-paced

Instructor paced courses follow a set schedule. The course team sets specific due dates for graded assignments and exams, and students complete the course within a defined time period. The course materials in the instructor-paced courses are released at specific times as the course progresses. An instructor pace course generally has multiple deliverable spreads throughout the courses. Students need to complete these deliverable with the passing grade to be granted a certificate.

self-paced

self-paced courses do not follow a set schedule. All of the course materials are released as soon as the course starts. Students can thus watch videos and practice problems through the course at their own speed.

The course team could structure self-paced courses in multiple ways. A popular format is one where assignments and exams do not have due dates so that students can complete the course at their own speed. In this format, students are issued certificates as soon as they have a passing grade in the course.

Another approach is to have at least one deliverable with a fixed due-date and issue certificates to students with a passing grade at the same time. Students could still complete the other course content at their own pace.

2.2.4 Related Work on Pacing

Nesterko and colleague [54] examined the relationship between the use of strict due date in MOOCs and course completion rates. They found that stricter due dates are associated with a 2% increase in certificate attainment rates. However, the ten courses they examined differed in content, design, student population, and many other dimensions. So while they saw an increase in completion in courses with strict dates, the author cautioned readers not to assert any causal link from this study. Along similar lines, Topolovec [70] explored the link between course delivery mode (student- paced or instructor-paced) and learners success. She also found the self-paced course to be correlated with a lower completion rate. And while she examined 35 different courses, these courses differed in content, course design, length, and population demographics. The author also cautioned the reader not to draw any causal relationship between course mode and the success, instead suggesting that more study is required to understand this relationship.

Outside MOOCs, a study of the traditional online course by France-Harris and others [67] and others compared the self-paced and instructor-paced delivery mode of a specific course. They found the self-paced course to be equally, or even more effective than the instructor-paced course. However, their experimental design had multiple limitations. First, the experiment was conducted on only 96 students in a controlled environment, a setup which differs from authentic MOOCs environment. Would such results hold in an actual MOOCs? Secondly, the author modified the self-paced course by incorporating design changes and by setting a minimum GPA requirement to enroll in the course. The GPA requirement runs counter to MOOC stated goals. Would we have seen the same results had both the self-paced and instructor paced course remained the same?

Reich and Mullaney [52] compared consecutive run of the same course with different pacing, self-paced, and instructor paced. They found few differences and non-significant differences in persistence, participation, and completion between the two runs. This study is the first one that established a causal relationship between pacing and learner outcome. However, the study does have limitations. First, the course they examined in the study, HerosX, is a non-technical course with an inadequate assessment structure. Are the results applicable in a quantitative course with a more comprehensive assessment structure? Secondly, the study examined only one possible outcome, certificate attainment. How does the self-paced course impacts other metrics and learning in general?

In summary, course pacing is one of a fundamental design choice that instructors have to make while creating MOOCs. The research literature on the impact of course pacing on learner outcome is limited. This makes it an interesting problem to explore with our framework.

Chapter 3

Framework to evaluate course design in MOOCs

The goal of evaluating course design is to determine the effect the course has on learners. Inspired by Cronbach's chapter on course evaluation from the book "Evaluation in Education and Human Services," we define the following two broad principles that define course evaluation [31]:

Outcomes of instruction are multidimensional, and a thorough evaluation needs to map out the effects of the course along these dimensions independently. Separately, each of these dimensions is narrow, and only by looking at them collectively, robust inferences about learning can be made.

A thorough evaluation needs to identify the aspects of the course where revision is desirable. This requires a broad understanding of the learners, the content of the course, and most importantly, learners' interaction with the course content. Only by understanding these behaviors, one can make robust inferences on what can be done to improve the course.

We built on the work done by Cronbach on course evaluation in a traditional setting and expanded it to a MOOC environment [31]. Our framework consist of the following three components:

- Process Measures

- Proficiency Measures
- Perception Measures

3.1 Process Measures

"Process Measures" examines events taking place in the course. Specifically, it studies student behavior and their interactions during instruction. The ability of MOOCs to capture fine-grained data provides an exciting opportunity to explore the instructional process at detail and depth never seen before. Process Measures holds special importance in the course evaluation process as it can provide key insights on how to improve the course.

We recommend answering the following questions to understand the student behavior and their interactions during instruction:

- **Course components:** How do learners engage with course components (Videos, problems and forums) with and without the change?
- **Course contents:** How do learners engage with specific course content (lectures, recitation, quick questions, practice problems, etc.) with and without the change?
- **High-achievers:** How does the behavior of "high-achievers" and "low-achievers" differ across the change?
- **Learners' performance:** What is the relationship between learners' grade and the way they engage with the course? How has this altered with the change in course design?
- **Completion:** How do completers and dropouts behave before and after the change?
- **Learners' pathway:** How does learner's behave over time (weeks)? How has this behavior evolved with the change in course design?

3.2 Proficiency Measures

"Proficiency Measures" examines changes observed in the learners. Existing proficiency measures in MOOCs such as completion and clicking are limited and do not support inferences about learning. A broader approach is required that goes beyond the current monolithic view of completion and takes into consideration the varying needs of the diversion population of MOOCs. To address this challenge, we defined an engagement score that captures students' engagement in MOOCs across multiple dimensions. We present details of it in the later section of this chapter.

We recommend answering the following questions to understand the impact of course design on students' success:

- **Definition:**What are the success metrics to track?
- **Impact:**What is the impact of course design on students' success metrics?
- **Sub-populations:** Is the difference in students 'success consistent across all students, or does it vary across sub-populations? Which population is impacted the most?

3.3 Perception Measures

"Perception Measures" examines students increase in interest in the subject, interpersonal outcomes (e.g. cooperative abilities), intrapersonal outcomes (e.g. self- understanding), and other broad course outcomes" [12, 38]. Even though we have taken a broad definition of the success metrics, none of the quantitative metrics can truly tell us what motivates learners and what happens in their head during the course. Student perceptions of learning have also been found to be much highly correlated with the overall ratings of teaching effectiveness [64, 11].

We recommend answering the following questions to capture student perception of the change:

- **Interest:** Does the course design increases or decreases students interest in the subject?

- **interpersonal:** How are the interpersonal outcomes (e.g., cooperative abilities) impacted by the change?

- **intrapersonal:** How are the intrapersonal outcomes (e.g., self- understanding) impacted by the change?

- **course components:** Does the student perception of the course and its associated components differs before and after the change?

3.4 Analytical Methods Employed

In this section, we summarize the methods employed to perform analysis in each of the sections discussed above.

3.4.1 Data Acquisition and Integration

We integrate clickstream data with a demographic, survey, and course metadata to create a 360-degree view of students' interaction with the course. The activity data, captured through clickstream in MOOCs, provides a record of learner activity of unprecedented scale and resolution. MOOCs have a diverse student body, and the demographic data enables us to identify the groups with specific demographics and socio-cultural traits. Survey data provide us insight on student attitudes and perceptions in the course. Finally, course metadata gives us information on the course structure and instructional materials. While each of these sources is valuable, the real values come when these disparate sources of data are combined to create a trace of students' interaction during the course across many dimensions.

3.4.2 Identifying sub-populations

We decomposed the overall learner population to sub-groups that behaves similarly or responds to changes similarly. In general, our philosophy has been to deconstruct the large student body to smaller groups whenever possible and measure the effect of treatment on each of the groups separately. We rely on unsupervised machine learning methods to identify clusters and causal inference methods to compare the effect of treatment on a group.

3.4.3 Static and Dynamic Analysis

We performed both static and temporal analysis on the data to examine student interaction between different courses and during the duration of the course. Temporal analysis can expose students pattern and behavior over time, giving us more fine-grained and insightful information on how to design course chapters.

3.4.4 Causal and Predictive Methods

We performed backward-looking (causal) and forward-looking (predictive) analysis to understand what causes students to behave in a certain way and how this behavior is likely to impact student success. Causal inference enables us to establish a causal link between a design change and student success while predictive models can provide more explanation on the mechanism behind the outcome of interest. The latter is also important as it can enable MOOC providers to provide target interventions or personalized delivery to students.

3.5 Defining Student success

We define student success in terms of the following metrics: completion, pass rate, and engagement.

3.5.1 Completion

A student completes the course if they attempt the final exam at the end of the course. A related metric is a dropout, which captures students that drop out of the course before the final exam.

3.5.2 Pass rate

A student passes the course if it achieves some pre-determined grade threshold for the course (60% in our case).

3.5.3 Engagement

A lot of research has been done on student engagement, but to date, the term is still not well defined. Kuh[39] , Robinson and Hillinger[61] and Handelsman, Briggs, and Towlers[27] have made important contributions in defining student engagement, although the focus of their effort has been on the traditional classroom setting. Dixon[19] expanded the concept of student engagement to online learning, and we use her definition in our study. In Dixon words, "A student is engaged if they use time and energy to learn the material, demonstrate that learning, interact with others in the class and become emotionally involved with their learning." Dixon [19] developed a Student Engagement Scale (OSE) and demonstrated its reliability and validity by performing multiple pilot study. In the same study, the OSE scale was shown to significantly and positively correlate with online student behavior as tracked by course management software.

We expanded on the OSE scale and linked it with actual online student activity to create a new quantitative score. We call this score engagement score (ES). The engagement score is a mix of the following three components: Academic Engagement, Cognitive Engagement, Social Engagement. The engagement index is a mean of these components.

Academic Engagement:

A student is academically engaged score if he or she participate in academic activities throughout the course. We use course videos as a proxy to measure academic engagement. Previous research on academic engagement has shown video views or assignment submissions to be an effective measure in measuring engagement [36]. We avoid using assignment submissions in our measurement of academic engagement as the version of the course understudy doesn't have any mandatory assignments. We do, however, recommend that assignments to be included in academic engagement if course do have required graded assignments. The equation 3.1 shows our formulation of academic engagement.

$$Academic\ Engagement = \frac{unique\ videos\ watched}{total\ required\ videos} \tag{3.1}$$

Cognitive Engagement:

A student is cognitively engaged if he or she make efforts to understand what is being taught, internalize learning and effectively demonstrate that learning. Unlike academic engagement, it is difficult to measure cognitive engagement directly. We rely on multiple proxies to infer student's cognitive engagement within the course. We use assessments to infer if a student is indeed learning [57]. We combine this with video pause events and rewatching pattern as previous research has shown to provide evidence that these events indicate a higher level of cognitive engagement [3, 66]. The equation 3.3 shows our formulation of cognitive engagement.

$$\begin{aligned} Cognitive\ Engagement = & \frac{problems\ attempted}{\theta_p * total\ problems} \\ & + \frac{videos\ rewatched}{\theta_r * total\ videos} \\ & + \frac{videos\ pause\ event}{\theta_u * total\ videos\ watched} \end{aligned} \tag{3.2}$$

where θ_p, θ_r and θ_u are thresholds that are determined experimentally by taking into consideration previous course distributions of these parameters and instructors expectation.

Social Engagement:

A student is socially engaged if he or she interact with others in the course, in a meaning full way. We operationalize this by using discussion forums as a proxy for student's interaction in the course.

$$\begin{aligned} Social\ Engagement = & \frac{forum\ active\ events}{\theta_{fa}} \\ & + \frac{forum\ passive\ events}{\theta_{fp}} + \\ & + \frac{forum\ post\ length}{\theta_{fl}} \end{aligned} \tag{3.3}$$

Where forum active events include the sum of all click events that are generated when a user posts or like a comment. On similar lines, passive events include all events that get generated when a user access a forum, searches a forum, or read a post. Finally, θ_{fa}, θ_{fp} and θ_{fl} are the thresholds that are determined experimentally by taking into consideration previous course distributions and instructors expectation.

3.6 Data-driven Framework to evaluate course design changes

In this section, we present a table that summarizes the key components of our Framework.

Table 3.1: Data-driven Framework to evaluate design changes in MOOCs

Topic	**Analysis**	**Method**	**Sources of Data**
Process Measure	1. How do learners engage with course components (Videos/problems/forums) with and without the change? 2. How do learners engage with specific course content (lectures, recitation, quick questions, practice problems, etc.) with and without the change? 3. How does the behavior of "high-achievers" and "low-achievers" differ across the change? 4. What is the relationship between learners' grade and the way they engage with the course? How has this altered with the change in course design? 5. How do completers and dropouts behave before and after the change? 6. How does learner's behaviors over time (weeks)? How has this behavior evolved with the change n course design?"	-Unsupervised learning to identify sub-populations -Statistical testing to compare changes -Time series analysis to capture interaction over time	clickstream, course metadata, achievement data

Topic	Analysis	Method	Sources of Data
Proficiency Measure	1. What are the success metrics to track? 2. What is the impact of a course design change on students success metrics? 3. Is the difference in students success consistent across all students, or does it vary across sub-populations? Which population is impacted the most?	-Unsupervised learning to identify sub-populations -Causal inference to measure the effect of treatment on success metrics -Supervised learning to understand the mechanism behind the outcome of interest	clickstream, course metadata, achievement data
Perception Measure	1. Does the course design change increases or decreases students interest in the subject? 2. How are the interpersonal outcomes (e.g., cooperative abilities) impacted by the change? 3. How are the intrapersonal outcomes (e.g., self- understanding) impacted by the change? 4. Does the student perception of the course and its associated components differ before and after the change?	-Statistical testing to compare changes	survey

Chapter 4

Case Study: Evaluating pacing in MOOCs

The course under study here is the first course (SC0x - Supply Chain Analytics) among a series of five courses that learners pursue to receive an MITx MicroMaster certificate in the supply chain (MITx MicroMasters® program in SCM) [50]. Each course, however, is independent and on successful completion of the course, learners are issued a course certificate. The first course in the Micromaster program is abbreviated as SC0x while the last course is SC4x (Supply Chain Technology and Systems), where the numbers represent the recommended sequence of the course. Collectively, they are known as SCx courses.

The first course, which we will call SC0x for the rest of the document, introduces learners to analytical tools and methods used in the supply chain. The course is designed to give both novices and specialists supply chain professional sufficient mathematical maturity to undertake future academic course work in the supply chain. The course is a recommended pre-requisite to all the subsequent course in the MIT supply chain series. SC0x, for obvious reasons, is of prime importance to the Micromaster teaching team as it offers an entry point into the program. It has the highest enrollment and unfortunately the highest dropout among all the other courses. The analytics of this course is well studied, and the previous course changes n late 2017 to improve the course content had led to reduced dropout and improved the pass rate.

SC0x was first offered in 2016 as a stop-gap to support existing SCx learners to get a solid foundation in mathematics (See figure 4-1 for history of SC0x). Previous to that, SC1x was the de facto entry point to the Micromaster program. Feedback from the learners and faculty pointed out to create a refresher course on mathematics for supply chain professionals, and thus SC0x was born. The first run of the course ran for a shorter duration, had a different structure, and targeted current learners in the Micromaster program. SC0x then went through a comprehensive redo, and the first full-fledged version was launched in early 2017. Another series of changes were made at the end of 2017 to refine the content and assessments. We don't use data from 2016 run of the SC0x as that version of the course differed both in its structure and intent.

The first five-run of the SC0x was instructor-paced. The courses followed a set schedule where lessons were released weekly, and assignments were due every week. There were eight graded assignments, distributed throughout the courses, in addition to a midterm in 6th week and a final in the week. The distribution of the grades was as follows: graded assignment (20%), midterm (35%), and final (45%). Learners who score 60% of above were granted course certificate.

In 2019, the pacing structure of SC0x was changed to self-paced. With this change, all of the course materials were released as soon as the course started. Learners were able to watch videos and practice problems through the course at their speed. Graded assignments and midterm exam were removed, and the only graded content kept was the final exam. The final exam contributed 100% to the course grade. The final exam had a fixed due-date, and learners with a passing grade in the final exam were issued course certificate.

In terms of content and sequence, the self-paced version of SC0x was identical to the instructor-paced version of SC0x. The duration of both the courses was also the same. The timing of the content release was the major difference between these two-course type, although other changes did emerge as a consequence of the change in timing.

The table 4.1 compares the self-paced and instructor-paced course in terms of timing and content. The content and duration are identical. Graded content in the instructor-paced course is distributed throughout the course while in self-paced, 100% of the grade is assigned to the final exam. Graded assignments from instructor-paced were converted to practice problems, and while the total number of problems remain the same, ungraded assessments have increased slightly in self-paced courses. The enrollment and verification deadlines were pushed further into the course. These changes are consequences of change in timing between the two courses, although we argue that one must control for these changes

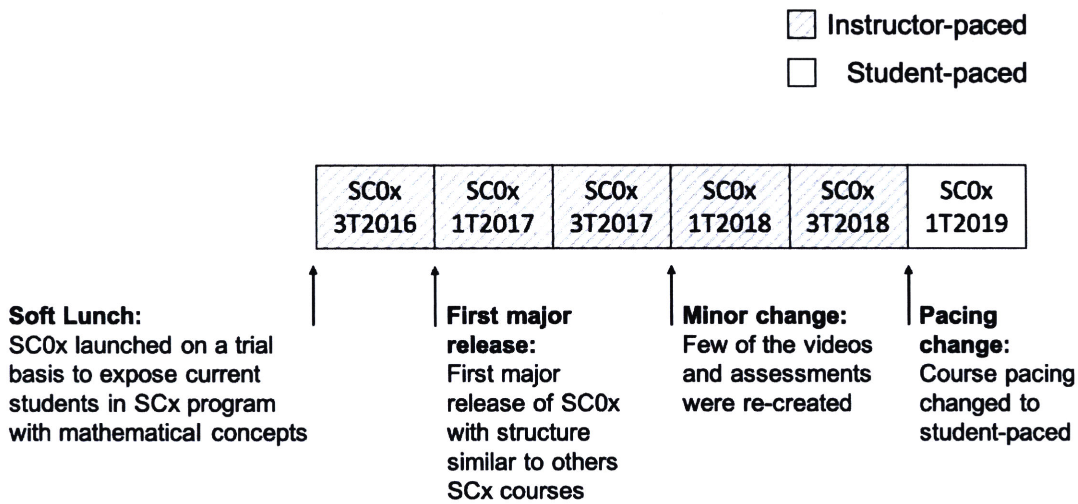


Figure 4-1: History of Supply Chain Analytics (SC0x) MOOC from 2016-present. SCOx was soft-launched in the third quarter (3T) of 2016 and formally launched in the first quarter (1T) of 2017. It remained an instructor-paced course until it transitioned to student-pacing in the first quarter (1T) of 2019.

Table 4.1: Comparison of course content across the versions of SC0x

	Instructor-Paced	**self-paced**
Content release	sequential	all at once
Number of runs	5	1
Total audit students	21434*	18607
Total verified students	1633*	1827
Verified conversion ratio	8%*	10%
Total weeks	12	12
Enrollment deadline	3 wks after course starts	9 wks after course starts
Verification deadline	4 wks after course starts	9 wks after course starts
Grade Distribution	Assignment (20%) Midterm (35%) Final (45%)	Final (100%)
Total hours of videos	1061*	1098
Total ungraded problems	376*	410
Total graded problems	102*	13

* Averaged over five runs

to measure the true impact of pacing on student success.

The table 4.2 compares the self-paced and instructor-paced course in terms of demographics (See Appendix for demographic changes over all previous runs of the course). A chi-squared test is performed on each dimension. The tables report the statistical and practical significance of the comparisons. The latter is reported in terms of phi-coefficients. By convention, phi>.7 is considered a large effect, phi>.05 large and phi>0.3 a medium effect.

We do observe some differences in the composition of demographics, but the effect size is weak, although significant. The most noticeable change is seen in mean age where a 25% increase is observed in learners in the age bracket from 30 to 40, and a similar percent decrease is seen in learners in the age bracket from 20 to 30. Why did this change happen? Without additional data, one can not present any statement on the reason for this shift. Our best explanation is that the self-paced course incentivized more working professionals to enroll in the course, thus leading to an increase in learners in the age bracket from 30 to 40. Apart from age, other demographics features remain mostly the same.

Thee table 4.3 compares the self-paced and instructor-paced in terms of registration and joining date. A noticeable change is observed in the verification deadline that is both statistically significant and has a large effect size as well. This change, however, is expected as by pushing the date further, we removed the students' incentive to pay for the course verification earlier. Can this negatively impact completion or pass rate? We explore this aspect in more details later.

Finally, we compare data from the entrance survey for the self-paced, and instructor paced course (Figure 4.4). We conditioned here on learners that have responded to the entrance survey (60%). These learners are more likely to complete and pass the course, and thus, care must be taken to generalize this result for the whole student body. The composition of responses remains quite the same, implying that employment structure and reason to enroll has not changed within the different types of courses.

The instructor and self-paced version of SC0x are similar along many dimensions and differ only by pacing. Few changes did emerge as a consequence of a shift to student-pacing: grade distribution, registration deadline, and some subtle demographic changes. If we can control for these changes, then the self-paced run of SC0x provides an exciting opportunity to evaluate pacing on MOOCs.

Table 4.2: Comparison of demographic features between across the versions of SC0x. A chi-square test is performed on each dimension. The table reports the statistical and practical significances of the comparison. The former is reported in terms of p-value and the later in terms of phi coefficient.

	Instructor	**Student**	ϕ	**p-value**
gender: female	22.7%	25.4%		
gender: male	77.1%	74.4%	0.064	0.023*
gender: others	0.2%	0.2%		
age: 0-20	0.3%	0.2%		
age: 20-30	36.9%	29.5%		
age: 30-40	44.7%	55.1%	0.211	0.000***
age: 40-50	13.3%	10.6%		
age: 50+	4.8%	4.7%		
edu: jr high school	0%	0%		
edu: high school	6%	5%		
edu: associate	2%	3%		
edu: bachelor	50%	52%	0.148	0.053
edu: masters	38%	37%		
edu: PhD	2%	2%		
edu: others	17%	16%		
country: US	32%	30%		
country: IN	11%	13%		
country: BR	5%	5%		
country: MX	3%	2%	0.087	0.033*
country: CN	3%	3%		
country: CA	3%	3%		
country: ES	3%	2%		
language: (25<x<50)	17%	16%		
language: (50<x<75)	41%	44%	0.046	0.138
language: (75<x<100)	42%	40%		
hdi: (x<0.5)	1%	0%		
hdi: (0.5<x<0.7)	21%	24%	0.094	0.008**
hdi: (0.7<x<0.9)	30%	29%		
hdi: (x>0.9)	49%	47%		
os: Windows	86.1%	88.9%		
os: Mac OS X	12.4%	10.2%	0.087	0.031*
os: Mobile OS	1.0%	0.6%		
os: Linux	0.5%	0.0%		

Note: ϕ is the phi coefficient for chi-square test. * p < .05 ** p < .01 *** p < .001

Table 4.3: Comparison of enrollment and verification timing across the versions of SC0x. A chi-square test is performed on each dimension. The table reports the statistical and practical significances of the comparison. The former is reported in terms of p-value and the later in terms of phi coefficient.

	Instructor	**Student**	ϕ	**p-value**
enrollment: 30 days before course	40%	36%	0.051	0.310
enrollment: 14 days before course	14%	12%		
enrollment: Before start of course	23%	22%		
enrollment: within 14 days of course start	18%	16%		
enrollment: within 30 days of course start	5%	5%		
enrollment: after 30 days of course start	0%	9%		
verification: 30 days before course	11%	18%	0.803	0.00***
verification: 14 days before course	6%	6%		
verification: Before start of course	14%	19%		
verification: within 14 days of course start	20%	20%		
verification: within 30 days of course start	48%	8%		
verification: after 30 days of start of course	1%	29%		
edx_join: Before 1 year	36%	38%	0.116	0.00***
edx_join: Before 1 month	36%	35%		
edx_join: Before start of course	19%	15%		
edx_join: After start of course	9%	11%		

Note: ϕ is the phi coefficient for chi-square test. * p < .05 ** p < .01 *** p < .001

Table 4.4: Comparison of entrance survey features across the versions of SC0x. A chi-square test is performed on each dimension. The table reports the statistical and practical significances of the comparison. The former is reported in terms of p-value and the later in terms of phi coefficient.

	Instructor	Student	ϕ	p-value
employment: Unemployed	9%	9%		
employment: Employed	78%	80%		
employment: Full-time student	10%	8%	0.042	0.498
employment: Retired	0%	0%		
familiarity: Not at all familiar	6%	4%		
familiarity: Slightly familiar	16%	14%		
familiarity: Somewhat familiar	48%	49%	0.075	0.033*
familiarity: Very familiar	24%	27%		
familiarity: Extremely familiar	5%	6%		
language: Weak	0%	0%		
language: Basic	0%	1%		
language: Intermediate	11%	11%	0.055	0.270
language: Fluent	58%	57%		
language: Proficient	31%	31%		

Note: ϕ is the phi coefficient for chi-square test. * $p < .05$ ** $p < .01$ *** $p < .001$

Chapter 5

Data Collection and Feature Engineering

In this chapter, we provide our approach for data collection. We begin by presenting details on the course structure and content. Later, we share sources of data and the list of covariates extracted from the data.

5.1 Course Structure and Content

5.1.1 Course Structure

All of the SCx courses consist of twelve weeks: a welcome week, eight content-based weeks, two prep-weeks, one Midterm exam, and one Final exam [49]. Each week is released on Wednesdays at 15:00 UTC. Graded assignments and exams are also always due at 15:00 UTC on the specified day.

Welcome Week

Week 0 is Welcome Week, and it provides information on course logistics, software, and background material that should help students prepare for the course.

Content-based Week

Prep Week: Week 5 and 11 are the prep-weeks. These weeks are meant to give breathing space to students to prepare from the midterm and final exams. No new graded assignments or new contents are released in these weeks.

Exam Week

Week 6 and 12 will consist of a midterm and final exam, respectively. The midterm exam is worth 35% of the total grade. The final exam is worth 45

5.1.2 Course Contents

Lecture Videos

Each content-based week usually has two lessons. Each lesson consists of a bundle of short videos interspersed with Quick Questions and discussions. Each video is between 2 to 12 minutes, and a whole lesson ranges from 40 to 60 minutes.

Recitations Videos

These are the videos that explain how to solve a problem using a specific method. Recitations are optional but highly recommended.

Quick Questions

Quick Questions (also known as finger exercises) are optional and ungraded questions interspersed in lecture videos that teach the content discussed in the video.

Practice Problems

Practice Problems are ungraded problems that enable students to practice and test their skills on concepts covered in that week. Each problem has multiple parts, and students are encouraged to view the detailed explanation after they have attempted the problem. Like quick questions, these problems are not graded and are optional.

Supplemental Material

These materials are available for verified MicroMasters students and are will also have access to supplemental videos, documents, and practice problems. These materials are intended to round out the supply chain education of the student - but are not covered in the Graded Assignments or Final Exam.

Graded Assignments

There are eight weekly Graded Assignments, and they are worth 20% of the total grade. Each content-based week contains one graded assignment section. Graded assignments are due two weeks after the week is released.

Midterm Exam

There is one midterm exam, and it is worth 35% of the total grade. The midterm exam is released at the start of week 6 and due after a week. Unlike graded assignments, a midterm exam has a time limit associated with it and must be completed within the allowed duration (generally 3 hours).

Final Exam

The final exam is released in the last week of the course (week 12) and is worth 45% of the total grade. Like the midterm exam, it is made available for only one week. Also similar to midterm, the final exam is timed.

5.1.3 Types of Learners

Learners who enroll in the course can be classified into two groups: Verified and Audit. Verified learners pay for the course, can access graded course content and are issued course certificate on successful completion of the course. Verified learners also have access to supplemental material, which contains additional and optional information on the subject. Audit learners, as the name suggest, are in the course to audit the course. They don't pay any course free and are not issued any certificate at the completion of the course. Audit learners have access to all course videos and most ungraded problems(except the problems in the supplemental material) but can't access any of the graded content in the course.

5.2 Data Systems and Sources

5.2.1 Learner Demographics Data

Information about learner demographics is extracted from the course registration page. This includes traditional demographics data as well as static user data related to the online environment such as user device, user operation system, registration date, etc. The inclusion of traditional static learner data remains vital, given the well-established correlations that exist between this data and student success.

5.2.2 Survey Data

Pre- and post-course surveys are used to get insights into student attitudes and perceptions. Both these surveys reveal greater insight on student's motivation and their intent to join the course. However, these surveys are optional and self-recorded, and therefore, the data from these surveys is often limited and is subject to various response biases. Specifically, We noticed that the response rate for exit surveys are really low as a large part of the learners have dropped out of the course before reaching to that stage. Nevertheless, we believe that ours is one of a few studies where data from the survey has been used in conjunction with online course activity data to develop student success models and intervention analysis methods.

Table 5.1: List of demographic features extracted from the platform

	Name	Description
u_1	user.gender	gender, as specified by the student
u_2	user.age	age, as specified by the student
u_3	user.level_of_education	level of education, as specified by the student
u_4	user.country	country of residence, as specified by the student
u_5	user.date_joined	date when the student joined edx.com
u_6	user.last_login	date when the student last accessed the course webpage
u_7	user.enrollment_date	date when the student enrolled in the course
u_8	user.verification_date	date when the student changed enrollment status from "audit" to "verified"
u_9	user.english_index	English Proficiency Index of the country (u4)
u_{10}	user.hdi	Human Development Index of the country (u4)
u_{11}	user.browser.median	median type of browser used by the student to access the course webpage
u_{12}	user.os.median	median type of os used by the student to access the course webpage

Table 5.2: List of features extracted from entrance survey

	Name	Description
s_1	survey.ent_responded	Whether the student has responded to survey or not
s_2	survey.career	Whether the student enrolled in the course to advance career or not
s_3	survey.education	Whether the student enrolled in the course to advance education or not
s_4	survey.fun	Whether the student enrolled in the course to enjoy or not
s_5	survey.familiarity	Familiarity of the student with the topics in this course
s_6	survey.utility	Importance student gives to learning the materials in this course
s_7	survey.intented_hours	Number of hours student intends to spend on this course per week
s_8	survey.online_courses	Number of online courses student has completed in the past
s_9	survey.intent_verified	Whether the student intends to earn the certificate or not
s_{10}	survey.intent_assess	Number of assignments student intends to complete
s_{11}	survey.fluency	English language fluency of the student
s_{12}	survey.empstatus	Whether the student is employed or not
s_{13}	survey.educ_pts	Highest educational level of the student's parents

Table 5.3: List of features extracted from exit survey

	Name	Description
s_{14}	survey.exit_responded	Whether the student has responded to survey or not
s_{15}	survey.expect	Whether the course met student expectation or not? (5 point scale)
s_{16}	survey.learn	Whether the course leaned from the course or not? (5 point scale)
s_{17}	survey.difficulty	The difficulty level of the course (5 point scale)
s_{18}	survey.course_quality	Quality of the instruction in the course (5 point scale)
s_{19}	survey.platform_quality	Quality of the software platform in the course (5 point scale)
s_{20}	survey.recommend	Whether the student will recommend this course to another student? (5 point scale)
s_{21}	survey.future_course	Whether the student will enroll in a future MITx course? (5 point scale)
s_{22}	survey.videos_quality	Quality of the videos in the course (5 point scale)
s_{23}	survey.questions_quality	Quality of the questions in the course (5 point scale)
s_{24}	survey.hw_quality	Quality of the homework in the course (5 point scale)
s_{25}	survey.forum_quality	Quality of the forums in the course (5 point scale)
s_{26}	survey.discuss	Importance of discussion forum in learning (5 point scale)
s_{27}	survey.perc_videos	Percentage of videos student interacted with in the course
s_{28}	survey.perc_hw	Percentage of homework student interacted within the course
s_{29}	survey.perc_exams	Percentage of exams interacted with in the course
s_{30}	survey.hours	Average number of hours spend per week
s_{31}	survey.max_hours	Maximum number of hours spend per week
s_{32}	survey.cert	Whether the student wanted to earn a certificate or not?
s_{33}	survey.soc_comm	Whether the student felt connected with the course community (5 point scale)
s_{34}	survey.soc_teach	Accessibility of teaching team in the course (5 point scale)
s_{35}	survey.soc_future	Whether the student intends to stay in contact with the person met in the course

5.2.3 Activity Data

Activity data is captured using clickstream. A clickstream is a record of a user's activity on the online course platform and includes interactions such as page views, video play/pause/skips, problem submission, forum read, etc. A record gets generated for every user interaction with the host server, and the resulting data is stored in JavaScript Object Notation (JSON) format. We found clickstream logs to be very rich in content, but most of them are raw and are highly-granular. Add to this, the massive scale and size of these logs, and it becomes a challenge to use them in any analysis task. We built on the work done on feature engineering of clickstream exports by Boorks[8] and Veeramachaneni[71] and developed a modular and flexible aggregation method to generate interpretable, meaningful and predictively features quickly and iteratively.

5.2.4 Course metadata

Detailed information about the course structure and instructional materials are extracted from the edX platform and linked with clickstream export to generate content specific features. This includes information about video lectures (type, section name, release date), forum(chapter name) and assignments (type, graded, release data). Linking this metadata to course activity enabled us to group features by content type, which reveals useful insight into how students approach different course contents.

5.2.5 Achievement data

Achievement data includes grades associated with examinations, assignments, and other graded content. This data connects student success in graded problems with course-specific objectives. So while one can extract individual grades of problems from activity data, the achievement data contextualize these grades, such as through weighting and retry opportunities.

Table 5.4: List of general course features extracted from the platform

	Name	Description
c_1	course.online_time	Total time spent on all resources
c_2	course.sessions	Number of web sessions
c_3	course.progress	Number of progress pages views
c_4	course.active_days	Number of active days

Table 5.5: List of video features extracted from the platform

	Name	Description
v_1	video.unique_count	Number of distinct video lectures
v_2	video.duration	Total time spent on lecture resources
v_3	video.pct_completed	Percentage of video lectures completed ($v_1/total\ videos$)
v_4	video.rewatch_count	Number of lectures rewatched
v_5	video.load_video	Number of video load events
v_6	video.play_video	Number of video play events
v_7	video.pause_video	Number of video pause events
v_8	video.seek_video	Number of seek video events
v_9	video.speed_change	Number of speed change events
v_{10}	video.show_transcript	Number of transcript events
v_{11}	video.time_to_watch	Average time from the release of the video to the moment student played the video

5.3 Summary Statistics

The table 5.11 presentation summary statistics of the last five runs of Supply Chain Analytics (SC0x) in terms of demographics. The gender composition over the last five runs has remained the same, with 77 percent of the learners specifying them as male and the others as female. We break down age into five buckets to get a better idea of the distribution of age across the course. Most learners in SC0x lie in the age bracket of 30 to 40, followed by learners in the age bracket from 20 to 30. We do see some variation in age over the last five runs of the course, especially in the top two represented age brackets. Most learners in SC0x had either an undergraduate degree or a graduate degree, confirming the graduate level focus of this program. The educational makeup has remained the same ever since

Table 5.6: List of forum features extracted from the platform

	Name	Description
f_1	forum.comment_created	Number of forum comments posts
f_2	forum.response_created	Number of forum response posts
f_3	forum.response_voted	Number of votes
f_4	forum.searched	Number of forum search queries
f_5	forum.comment_viewed	Number of forum comments views
f_6	forum.response_viewed	Number of forum response views
f_7	forum.length	Average length of forum posts
f_8	forum.duration	Total time spent on forum resources
f_9	forum.source	Course content (video/problem/forum) that bought user to the forum

Table 5.7: List of problem features extracted from the platform. These features were extracted for both graded and ungraded problems.

	Name	Description
p_{12}	prob.viewed	Number of problems views
p_{13}	prob.unique_submitted	Number of distinct problems submission
p_{14}	prob.submitted	Number of problems submission
p_{15}	prob.pct_completed	percentage of problems submitted
p_{16}	prob.problem_show	Number of problem show events where a problem show is generated when the answer to a problem is shown
p_{17}	prob.show_answer	Number of show answer events where a show answer is generated when a detailed explanation to the answer of the problem is shown
p_{18}	prob.success	Number of distinct correct problems
p_{19}	prob.duration	Total time spent on problem resources
p_{20}	prob. time_to_attempt	Time between release of problem and attempt
p_{21}	prob.avg_submit_prob	Average number of submissions per problem
p_{22}	prob.avg_submit_per_correct	Ratio of number of graded problems attempted to number of distinct correct problem
p_{22}	prob. pct_correct_submit	Ratio of number of correct graded problems to number of distinct graded problem
p_{23}	prob. first_attempt_correct	Ratio of number of correct problems on first attempt to number of distinct problem

Table 5.8: List of video metadata extracted from the platform

Name	Description
Type	Lecture, Recitation or Supplemental
Section Name	Module or section where the video is located
Release Date	Time when the video was released by the course team
Count	Count of all video resources grouped by type/section name
Duration	Duration of all video resources grouped by type/section name

Table 5.9: List of problem metadata extracted from the platform

Name	Description
Type	QQ, PP, GA, Supplemental, Midterm, Final
Section Name	Module or section where the problem is located
Release Date	Time when the problem was released by the course team
Graded	Whether the problem is graded or not
Duration	Count of all problem resources grouped by type/section name

Table 5.10: List of achievement data extracted from the platform

Name	Description
Absolute Grade	Absolute grade of the problem
Weighted Grade	Weighted grade of problem where we calculate the contribution of the problem to the overall course grade
Running Grade	Running weighted grade of the student

this course was launched. An overwhelming majority of the learners are from the United States (32%), followed by India (11%) and Brazil(5%). No substantial variation is seen in the distribution of the country of residence. The English index measures the average level of English language skills of the country of residence of the learners [21]. By convention, index>0.75 is considered highly proficient, index>0.5 as moderately proficient and index>0.25 as low proficient. Most learners come from countries where the English language skills are at least moderately proficient with the distribution of highly proficient and moderately proficient remaining the same over the runs of the courses. The HDI index tracks the human development index of the country of residence of the learners. By convention, index>0.75 is considered high HDI, index>0.5 as moderate HDI and index>0.25 as low proficient. Almost exclusively, all learners come from countries where the HDI is at least ranked as moderate. The highest represented group of learners are from countries with a very high HDI. Lastly, the most represented operating system is Windows (88%) followed by Mac OS (11%). The mobile users have been embarrassingly low, although the measurement only tracks the most used operating system by a learner and doesn't represent the distribution of time spent by learners on various devices. In summary, the composition of demographics has mostly remained the same over the various rune of SC0x. Having said that, we do need to carefully observe and possibly control for any demographic shift that could emerge from changes in course content.

The table 5.12 presents summary statistics of the last five runs of Supply Chain Analytics (SC0x) in terms of clickstream activities. The data is conditioned only on verified learners, but even then, we observe large variation in the clickstream activities, suggesting that there are groups with different intentions and patterns within the verified learners. As an example, let us look at video count, which tracks the number of unique videos viewed by the learners. The mean video count is 145 through the course. However, learners at 25% percentile only view 50 unique videos, while those at 75% percentile views 213 videos. The other video features follow the same pattern where we see a large variation between groups of learners at 25% percentile and 75% percentile. We have grouped forum activities into two classes: passive and active. Active events are generated when a learner posts or vote a comment. Passive events include events that get generated when learner searches, views, or read a comment. The table shows that the forum features are highly skewed with only a very small percentage of learners generating most of the forum activities. This is consistent with findings from other research [37]. The problem events are grouped into two classes: graded or ungraded. Similar to video features, we observe large variation even when the data is conditioned on verified learners, suggesting the possible existence of subgroups within the verified learners. This supports our initial claim that one needs to segment and understand these subgroups to better understand the impact of course changes within these subgroups.

Table 5.11: Summary Statistics of the demographic features for SC0x

	mean	std	min	max
gender: female	23%	1%	22%	25%
gender: male	77%	1%	74%	78%
gender: others	0%	0%	0%	0%
age: 0-20	0%	0%	0%	0%
age: 20-30	35%	4%	30%	39%
age: 30-40	47%	6%	41%	55%
age: 40-50	13%	2%	11%	15%
age: 50+	5%	1%	3%	6%
edu: jr high school	0%	0%	0%	0%
edu: high school	6%	1%	5%	7%
edu: associate	3%	0%	2%	3%
edu: bachelor	51%	1%	50%	52%
edu: masters	38%	1%	37%	39%
edu: PhD	2%	0%	1%	2%
edu: others	1%	0%	1%	1%
country: US	32%	1%	30%	33%
country: IN	11%	2%	10%	13%
country: BR	5%	1%	4%	6%
country: MX	3%	1%	2%	4%
country: CN	3%	1%	2%	4%
country: CA	3%	0%	2%	3%
country: EG	3%	1%	2%	3%
english: (25<x<50)	16%	2%	15%	19%
english: (50<x<75)	42%	2%	39%	44%
english: (75<x<100)	42%	2%	40%	44%
hdi: (x<0.5)	1%	1%	0%	2%
hdi: (0.5<x<0.7)	21%	2%	19%	24%
hdi: (0.7<x<0.9)	29%	2%	27%	31%
hdi: (x>0.9)	49%	2%	47%	52%
os: Windows	87%	1%	85%	89%
os: Mac OS X	12%	1%	10%	14%
os: Mobile OS	1%	0%	0%	2%
os: Linux	0%	0%	0%	1%

Table 5.12: Summary statistics of clickstream data for SC0x

	mean	**std**	**25%**	**50%**	**75%**
video.unique_count	145	108	50	146	213
video.duration (mins)	1257	1019	393	1170	1838
video.rewatch_count	77	97	12	45	104
video.pct_completed	0.76	0.32	0.60	0.92	1.00
video.avg_time_to_watch (days)	22	14	15	18	31
video.load_video	267	238	79	232	384
video.play_video	761	1067	120	477	1035
video.pause_video	470	773	73	286	610
video.seek_video	392	615	39	196	504
video.speed_change_video	14	39	0	2	12
video.show_transcript	12	75	0	0	2
forum.active_events	2	7	0	0	1
forum.passive_events	64	136	0	12	71
forum.duration (mins)	88	169	0	19	101
prob.graded.submitted	57	51	0	54	107
prob.graded.unique_submitted	43	38	0	40	85
prob.graded.success	37	34	0	32	72
prob.graded.duration (mins)	517	490	28	474	848
prob.graded.pct_completed	0.44	0.39	0.00	0.43	0.85
prob.graded.problem_check	115	103	0	110	214
prob.graded.problem_show	8	16	0	1	9
prob.graded.showanswer	8	16	0	1	9
prob.graded.avg_submit_problem	1.33	0.20	1.22	1.32	1.38
prob.graded.correct_first_attempt	28	27	0	21	55
prob.graded.pct_correct_submit	0.85	0.11	0.81	0.86	0.93
prob.graded.avg_time_to_attempt (days)	8	1	8	8	8
prob.ungraded.submitted	232	193	54	200	389
prob.ungraded.unique_submitted	144	117	34	124	246
prob.ungraded.success	114	98	25	92	190
prob.ungraded.duration (mins)	1056	1183	138	683	1630
prob.ungraded.pct_completed	0.40	0.32	0.09	0.35	0.68
prob.ungraded.problem_check	464	387	108	397	774
prob.ungraded.problem_show	75	103	6	38	101
prob.ungraded.showanswer	76	104	6	39	102
prob.ungraded.avg_submit_problem	1.63	0.32	1.43	1.60	1.74
prob.ungraded.correct_first_attempt	86	77	16	67	141
prob.ungraded.pct_correct_submit	0.79	0.16	0.73	0.80	0.89
prob.ungraded.avg_time_to_attempt (days)	13	8	8	9	12

Chapter 6

Process Measurement I: Static analysis of events taking place in MOOCs

In this chapter, we perform a static analysis of events that took place in the self-paced and instructor-paced course. We use "static" to refer to analysis where we are concerned only with the final state of the activity and not how or what path was taken to reach to that final state.

Specifically, we want to address the following questions:

- **Course components:** How do learners engage with course components (Videos, problems and forums) with and without the change?
- **Course contents:** How do learners engage with specific course content (lectures, recitation, quick questions, practice problems, etc.) with and without the change?
- **High-achievers:** How does the behavior of "high-achievers" and "low-achievers" differ across the change?
- **Learners' performance:** What is the relationship between learners' grade and the way they engage with the course? How has this altered with the change in course design?
- **Completion:** How do completers and dropouts behave before and after the change?

6.1 Course Activity and Engagement

First, we examine changes in course activity between self-paced and instructor-paced courses. We conduct this analysis across three fronts: videos, problems, and forums.

Video Activities

The table 6.1 shows a comparison of video features between student- and instructor-paced course. A one-way analysis of variance (ANOVA) is performed on each dimension. The table reports the statistical and practical significances of the comparison. The former is reported in terms of p-value and the later in terms of Cohen's d value. By convention, $d>0.8$ is considered a large effect, $d>0.5$ medium and $d>0.2$ small. Absolute effect sizes can be extracted from the averages in the table.

Video activity varies significantly between both groups with small to medium effect size. Learners in instructor-paced course watch videos at significantly higher rates than learners in a self-paced course ($p<0.001$, $d> 0.4$). A similar effect is observed with video rewatching behavior. Finer grained events such as play and pause also report the same effect. The only exception is "show_transcript" event that has increased slightly. Further analysis of this event shows that the increase is driven because of the availability of "Spanish" language in the latest run of the course. Prior to that, only "English" language was available in the transcript option. Overall, learners in instructor-paced course complete on average 15% more videos than learners in the self-paced course.

Assessment Activities

We focus here on ungraded problems as the only graded assessment in the self-paced course is the final exam. The table 6.2 shows a comparison of problem features between student- and instructor-paced course. Problem activity varies significantly between both groups, but the effect size is minimal across most dimensions. For example, the number of problem attempts between both groups is statistically significant, but the effect size is so tiny that one can claim that no noticeable difference is observed between the groups. A similar observation is observed across other problem events. One noticeable difference is the 'number of show answer' and 'the percentage of correct submissions per problem' events. Learners in the instructor-driven course are more likely to view explanation to the solution of the

Table 6.1: Comparison of video activities across the versions of SC0x. A one-way analysis of variance (ANOVA) is performed on each dimension. The table reports the statistical and practical significances of the comparison. The former is reported in terms of p-value and the later in terms of Cohen's d value.

	mean		d	p-value[1]
	Instructor	Student		
unique_count	156	107	-0.454	0.000***
rewatch_count	87	43	-0.459	0.000***
duration (mins)	1353	912	-0.440	0.000***
pct_completed	0.8	0.7	-0.377	0.000***
load_video	284	208	-0.320	0.000***
play_video	810	584	-0.212	0.000***
pause_video	496	378	-0.152	0.000***
seek_video	431	251	-0.295	0.000***
speed_change_video	15	12	-0.089	0.000***
show_transcript	11	14	0.038	0.142

[1] * $p < .05$ ** $p < .01$ *** $p < .001$

problem as compared with learners in a self-paced course. Interestingly, the number of correct submissions per problem increases in a self-paced course when compared with the instructor paced, although the effect size is very small. This confirms that learners across both groups have the same skill set. In summary, no major differences in ungraded problem activity are observed across both groups.

Forum Activities

We did not observe any noticeable difference in forum activities between instructor- and self-paced courses. While the difference is statistically significant, the effect size based on Cohen's d value is very small. In summary, forum activity across both groups remains mostly the same.

6.2 Grade and Engagement

Next, we investigate how a student's level of engagement and activity correlates with her final grade. Our goal is not to compare grade nor to predict grade from her activity but rather to gain an insight into the relationship between a student's grade and the way she engages with the course.

Table 6.2: Comparison of assessment activities across the versions of SC0x. A one-way analysis of variance (ANOVA) is performed on each dimension. The table reports the statistical and practical significances of the comparison. The former is reported in terms of p-value and the later in terms of Cohen's d value.

	mean		d	p-value[1]
	Instructor	Student		
problem_check	469	445	-0.063	0.017*
submitted	235	222	-0.065	0.014*
unique_submitted	145	139	-0.056	0.034*
success	115	111	-0.042	0.110
problem_show	80	83	-0.220	0.000***
showanswer	81	59	-0.218	0.000***
duration	1052	1073	0.018	0.493
pct_completed	0.4	0.3	-0.240	0.000***
avg_submit_problem	2	2	-0.079	0.000***
pct_correct_submit	0.8	0.8	0.125	0.000***
correct_first_attempt	87	82	-0.069	0.009**

[1] * $p < .05$ ** $p < .01$ *** $p < .001$

Table 6.3: Comparison of forum activities across the versions of SC0x. A one-way analysis of variance (ANOVA) is performed on each dimension. The table reports the statistical and practical significances of the comparison. The former is reported in terms of p-value and the later in terms of Cohen's d value.

	mean		d	p-value[1]
	Instructor	Student		
passive_events	62	74	0.089	0.000***
active_events	2	2	-0.039	0.133
duration (mins)	85	98	0.080	0.000***

[1] * $p < .05$ ** $p < .01$ *** $p < .001$

The figure 6-1 plots the median number of actions of a given type user take as a function of their final grade. Overall, we observe that the grade is generally proportional to their activity.

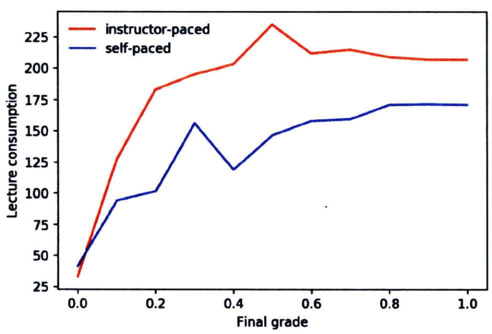
In both types of courses, the median number of problems submitted linearly increase with the student's final grade.

The grade increases non-linearly with a median number of videos watched. In the instructor-paced course, the grade increases monotonically with lecture consumption until 50%. Afterward, it falls somewhat and then remains constant after 60%. Perhaps this is due to SC0x being a highly quantitative course where those who have a strong quantitative background and have seen the content before have an easier time, while others seem to struggle. The self-paced course follows the same pattern but is shifted downwards by approximately 25%. Surprisingly, learners in a self-paced course can secure a perfect score by watching approximately 160 unique videos while watching these number of videos in an instructor-paced course will only get you a 20% grade. Perhaps this is due to learners starting late in the self-paced course and then making up by watching only the important videos. Similar behavior is observed with video rewatching where the curve straightens out early, and the gap between two groups widens to 50%. Nevertheless, both pattern demonstrates the importance of videos on final grade and also highlights that more video does not necessarily translate to a higher grade.

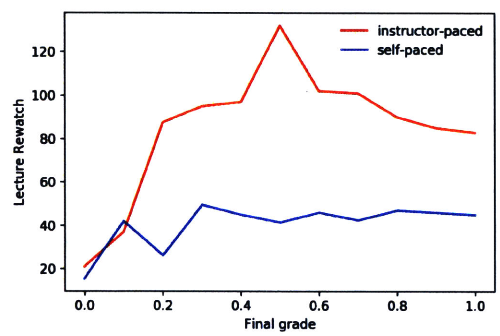
Finally, in both types of courses, the median number of forum activity linearly increase with the student's final grade. In the self-paced course, the curve is shifted upwards, implying that perfect scorers in that group view two times more posts than similar learners in the instructor-paced course. Perhaps the self-pacing nature of the course and the lack of instructional moderation pushes learners to explore more posts in a self-paced course.

6.3 Course Content-Type and Engagement

Next, we examine how students' interacted with various course content in both student- and instructor-paced format.

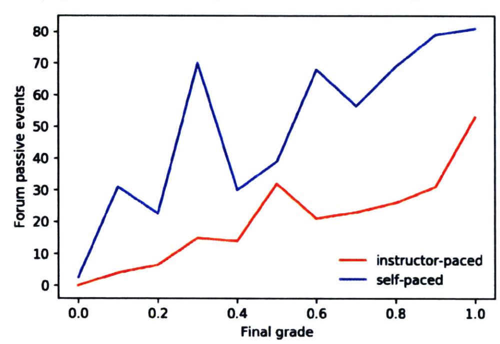


(a) Lecture consumption vs. grade



(b) Rewatching behaviour vs. grade

(c) Problem submissions vs. grade



(d) Course time vs. graded

Figure 6-1: Median count of actions of students with a given final grade in instructor-paced and self-paced version of SC0x.

Video Content

The figure 6-2 shows students' interaction with lectures and recitation. Readers are recommended to read the chapter on data collection to understand the difference between the types of videos. In a nutshell, lectures are mandatory while recitations are not. We did not observe any major difference in lecture viewing between both the groups. We do, however, see that more students in the instructor-paced format are likely to watch 100 percent of the recitation material as compared to learners in self-paced version. The decrease in overall video activity (as was shown in the previous section) is largely driven by recitation material and not by mandatory lectures.

Assessment Content

The figure 6-3a shows students' interaction with the three types of assessment content: practice problems, quick questions, and supplemental material. All of these assessments are ungraded. Readers are recommended to read the chapter on data collection to read more about these types. Surprisingly, no noticeable change is observed in quick questions and practice problems. Despite changing the pacing to self-paced, students were motivated enough to complete these ungraded questions. We do see some difference in the supplemental material. The change, however, is expected as supplemental materials are advertised to include content that is outside of the material covered in the final exam. Considering that many learners in self-paced format start the course very late, it is very natural to skip content that is marketed as outside the scope of the final exam.

The teaching team moved the graded assignment from the instructor-paced version and created two new sections in the self-paced format: Extra Practice Problems (Extra PP) and Exam Prep. We explore the interaction with these content types and their impact on the pass rate in figure 6-4. The figure only considers students that attempted the final. Students that pass the course attempted exam prep questions more than those who do not. The same trend is observed with Extra Practice Problems. Further, we notice a positive correlation between the score and the attempts in these assessment types. Notice that we do not claim that participation in these assessments causes the score to increase. We are just saying that students that passes the course are more likely to complete these assessment types. Figure 6-5 gives another viewpoint of problems type where we compare problem attempted during the duration of the course for both passed and failed students. The figure only considers students that attempted the final. The trajectory remains the same for both the student groups with the attempts peaking for all types in the last week. Students that pass the course attempts more problems and do these early in the course. That gives them time to spend more resource on Extra PP and Exam Prep during the end of the course. Students that fail the course attempts fewer problem during the early part of the course, forcing them to play catch-up game at the end. This leaves little time for them to attempt Extra PP and Exam Prep, which is reflected in the figure.

Temporal Analysis

Finally, we investigate students' interaction with course content during the duration of the course. The figure 6-6 visualizes topic consumption over time with the x-axis representing weeks and y-axis showing lecture consumption of a topic in minutes. In an instructor-paced format, videos are released every week, and the students are expected to view the videos within the next 14 days. The topic consumption follows this pattern with a clear, distinct spike every week. In a self-paced format, all the content is released all at once, and students can watch videos at their own pace. Unlike instructor-paced, there is no recurrent pattern across the weeks. The lecture consumption for module 1 and module 2 peaks very early in the course (week 1 and 3 respectively). There is no clear peak for module 3 with an almost equal proportion of students watching this module from week 4 to week 12. The lecture consumption of the other two modules peaks at the last week of the course. Learners took advantage of the flexibility offered to them in the self-paced version by pushing a good amount of work downstream. Also, most students did loosely watch modules in order, although many did not.

6.4 Performance and engagement

We now turn our attention to the activity of "high achievers," the students who secured A in the class. We investigate if "low achievers" has engaged differently in the instructor-paced version of the course versus the self-paced version [3].

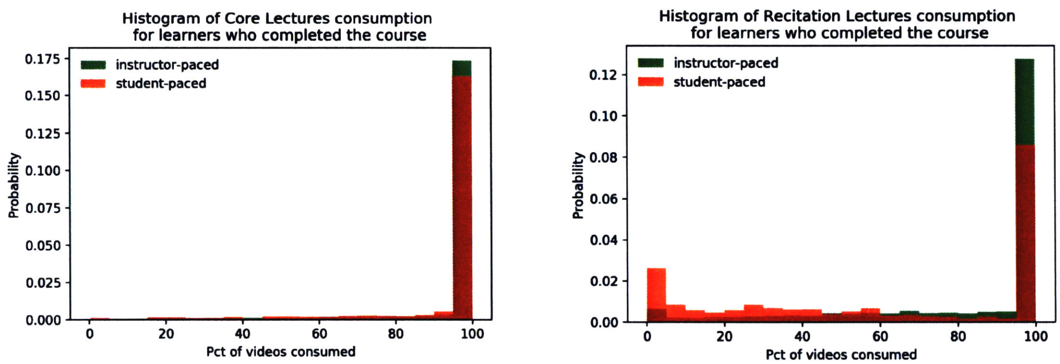
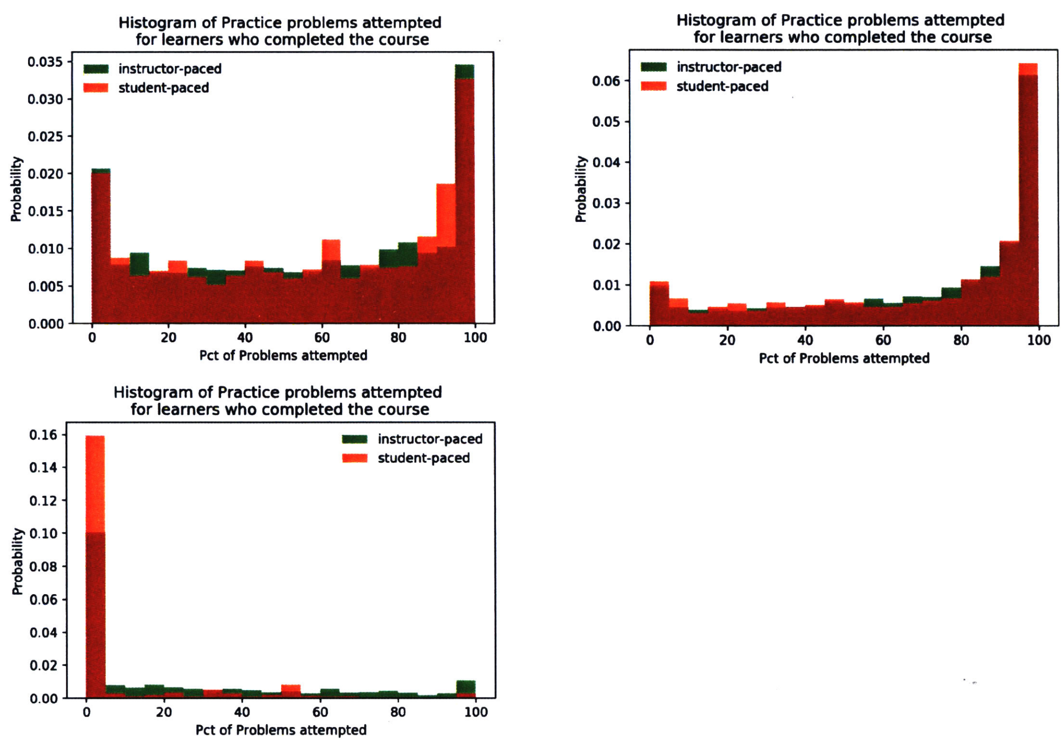


Figure 6-2: Probability distribution of core lectures (left) and recitation lectures (right) for completers in the instructor-paced and self-paced version of SC0x. Core lectures are required while recitations are optional.



(a) Probability distribution of quick questions (top left), practice problems (top right) and supplemental material (bottom left) for completers in the instructor-paced and self-paced version of SC0x. All of these assessments are ungraded and optional.

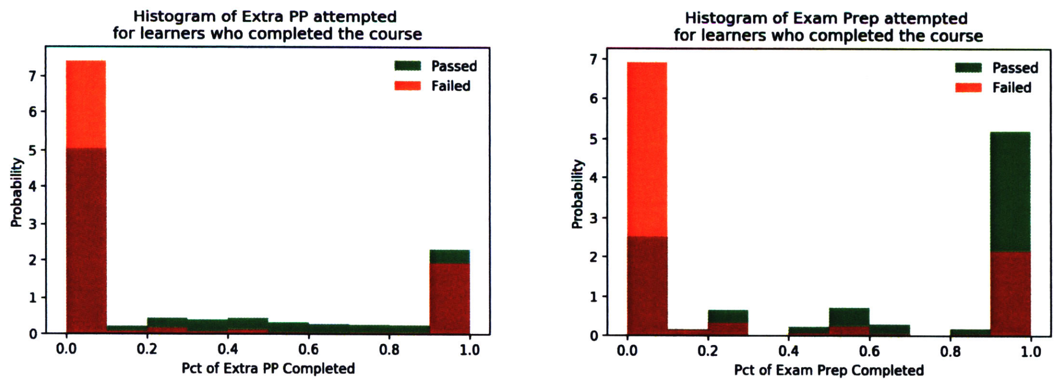


Figure 6-4: Probability distribution of Extra PP (left) and Exam Prep (right) for completers in the instructor-paced and self-paced version of SC0x. Both of these assessments are ungraded and optional. Extra PP is included at the end of module and Exam Prep is situated right before the Final Exam.

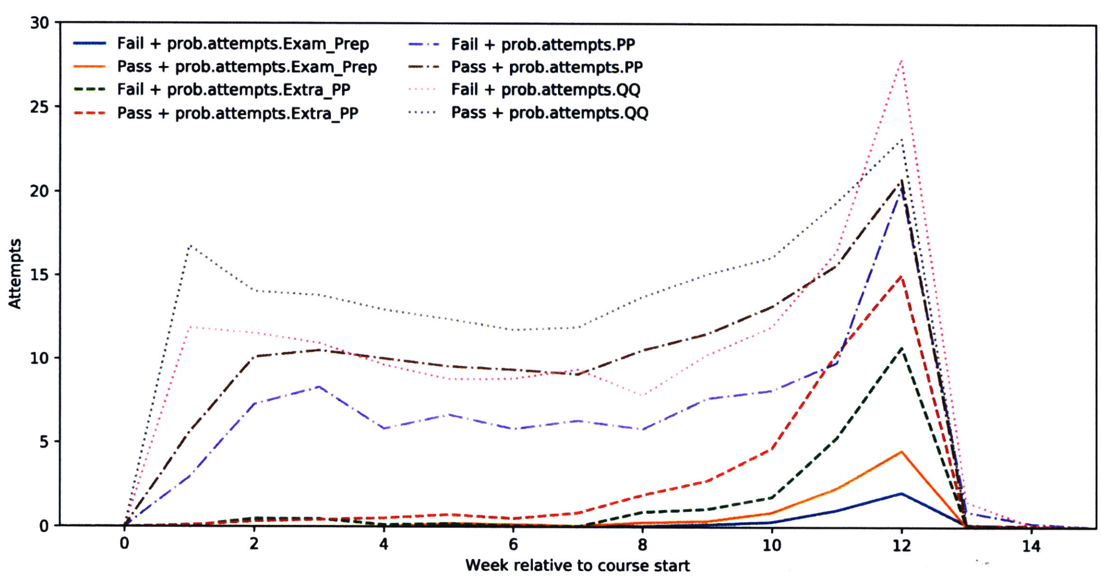


Figure 6-5: Engagement trajectory of problem activities separated by problem type for completers in self-paced version of SC0x. Y-axis represents number of problem attempted per week and X-axis shows the week relative to start of course.

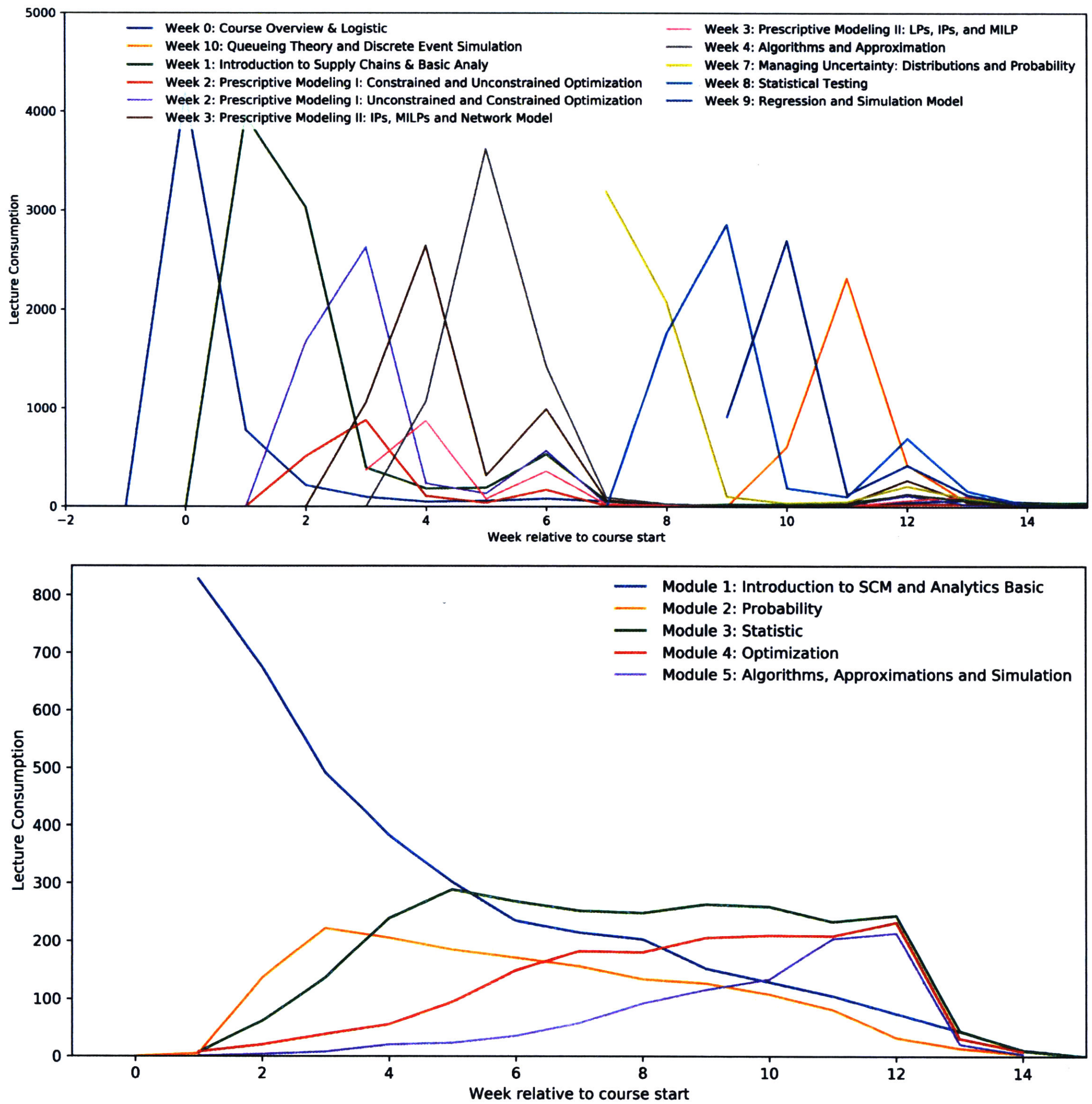


Figure 6-6: Engagement trajectory of video activities separated by chapter/module type for completers in instructor-paced (top) and self-paced version (bottom) of SC0x. Y-axis represents number of videos watched per week and X-axis shows the week relative to start of course.

Performance and Assessment Submission

The distinguishing attribute of high achievers is that they attempted more ungraded problems 6-7. The behavior is consistent across the courses with different pacing modes. Most of the high achievers attempted 80% of the ungraded problems. Surprisingly, there are some learners that attempted very few ungraded problems and were still able to secure A in the class. We investigated to check for the validation of the data for these learners and did not find any discrepancies. Our best guess is that these learners already had prior knowledge of the content of the course and were able to pass the exam without sufficient practice. A somewhat more practical insight comes from looking at the problem submissions trace for 'low achievers,' the students who completed the course but did not pass. Low achievers in instructor-paced course attempted relatively more problems than low achievers in the self-paced course, suggesting that an instructor-paced course was able to engage 'low achievers' for a far more longer time. We note this point and further explore it in the recommendation section.

Performance and Video Watching

The lecture watching differs between both groups on two accounts (6-7). First, we notice that 'high achievers' in instructor-paced version watch more videos than 'high achievers' in the self-paced course. This is in sync with the result of the previous section. Secondly, 'low achievers' in instructor-paced course end up watching almost the same amount of videos as high-achievers. The 'low-achievers' in the self-paced course, however, showed different behavior. A closer look at their distribution indicates that a subgroup within them did watch the same videos as the 'high achievers,' but there was a much larger subgroup that watched significantly fewer videos than both the 'high achievers' in the self-paced course and 'low achievers' in the instruction-paced course. This, again, highlights the negative impact on 'low achievers' in a self-paced course.

Performance and Forum Activities

The forum activities are consistent across both group to a larger extent (6-7). 'High achievers' in both groups participate more in forums than 'low achievers.' Unlike videos and problems, 'high achievers' in the self-paced course perform more activities in the forum when compared to 'high achievers' in the instructor-paced course.

6.5 Dropout and Engagement

We now investigate the engagement style in the context of dropout. Specifically, we are interested to understand if there are any changes in engagement style among dropouts between the self-paced and instructor-paced course.

Dropout and Video Activities

Figure 6-8 shows the histogram where the percentage of videos completed is shown on the x-axis and the probability of learners on the y-axis. The first thing that we notice is that the overall viewing pattern remains the same in both the course type. There are, however, some subtle differences. The distribution of the self-paced course is thickly spread towards zero, whereas an almost opposite phenomenon is observed in the instructor-paced course (Notice the visible green from 50% to 100%). On average, one can claim that learners who drop out of the self-paced course engage with the course at a lesser rate than the dropouts in the instructor-paced course. This aligns with what we have said before about the self-paced course. Learners at-risk of dropping out or not passing the course are more impacted due to the self-pacing nature of the self-paced course.

Dropout and Assessment Activities

Similar behavior is seen with ungraded problem submissions. Figure 6-9 shows the histogram where the percentage of problems completed is shown on the x-axis and the probability of learners on the y-axis. Similar to videos, learners in instructor-paced course engaged more with the problems then dropouts with the self-paced course. An interesting thing that pops out from this figure is a very high peak at zero for self-paced delivery, implying that dropouts in that format are twice more likely not to attempt any problem. Perhaps, deliverable with due dates distributed throughout the instructor-paced courses force more learners to engage in problems. In summary, we notice that dropouts in self-paced course engage less likely with those in instructor-paced version. We quantify this relationship in the subsequent sections.

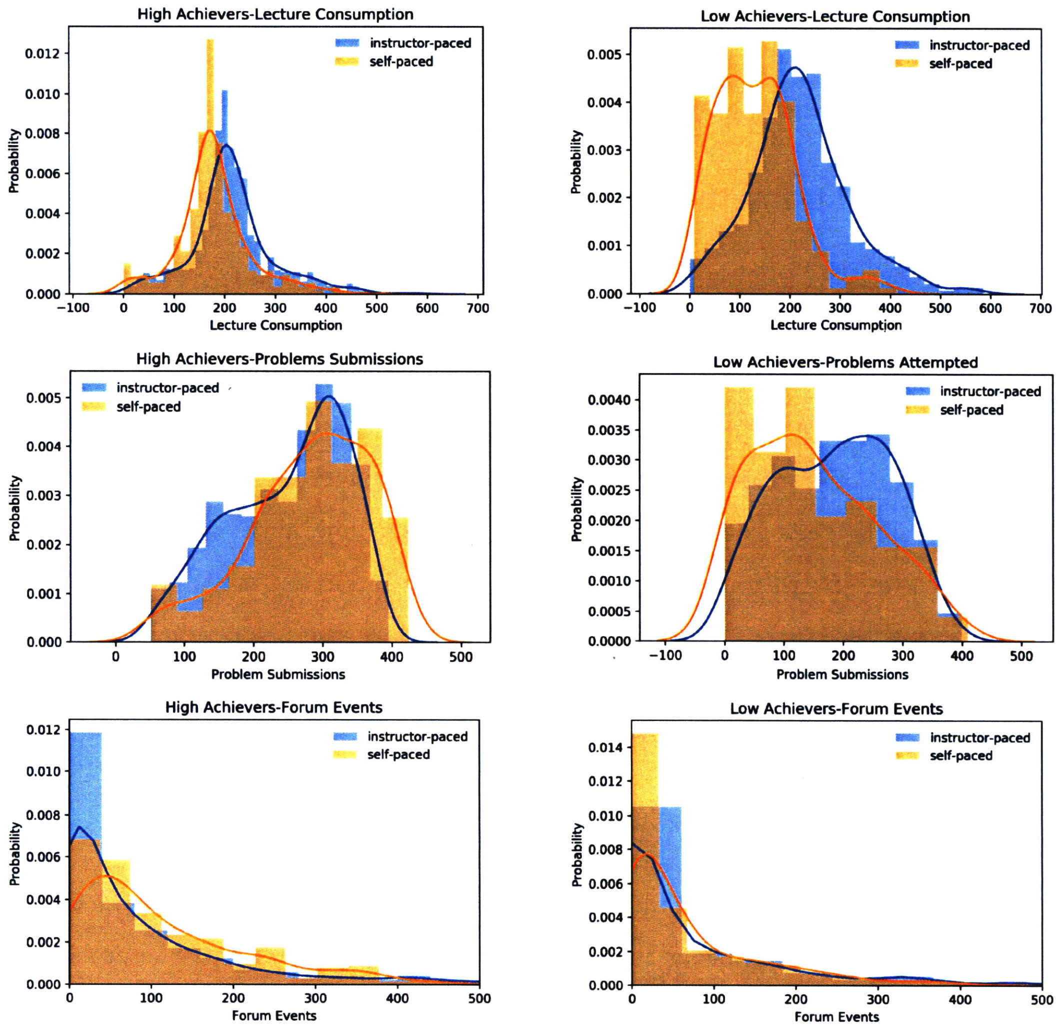


Figure 6-7: Probability distribution of Video, Assessment and Forum (top to bottom) activities of students with course grade of A (left) and F (right) in instructor-paced and self-paced version of SC0x. X-axis show the count of the particular activity for all the graphs.

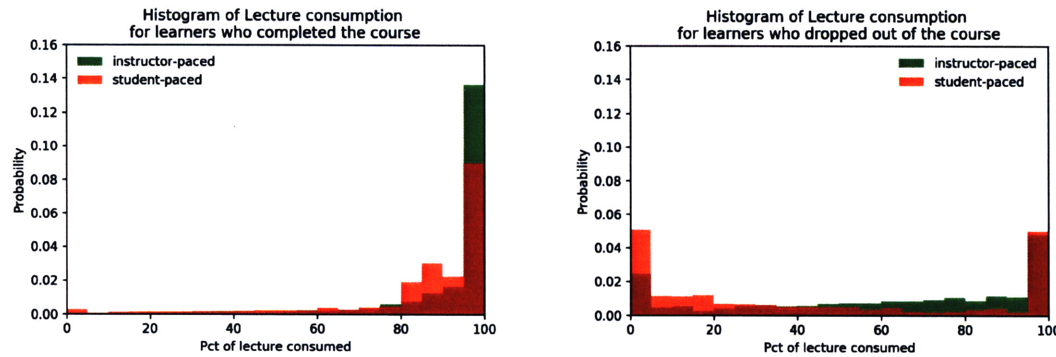


Figure 6-8: Probability distribution of percentage of videos consumed by completers (left) and dropouts (right) in instructor-paced and self-paced version of SC0x.

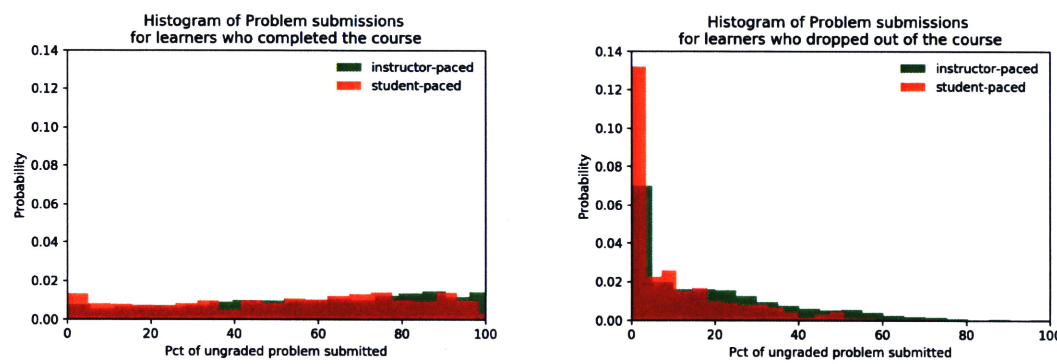


Figure 6-9: Probability distribution of percentage of ungraded problems submitted by completers (left) and dropouts (right) in instructor-paced and self-paced version of SC0x.

Chapter 7

Process Measurement II: Temporal analysis of events taking place in MOOCs

In this section, we present a temporal analysis of learners' engagement across both the self-paced and instructor-paced course. Specifically, we are interested in tracking learners' behaviors over time (weeks) and investigate if indeed a shift to the self-paced course has changed this behavior.

To address this, we developed a temporal classification method to identify a number of longitudinal engagement trajectories in MOOCs. Learners are classified based on their patterns of interaction with video lectures, assessments, and forums. We compare these clusters between the two versions of the course. Finally, we comment if user demographics dictate these trajectories and if following a certain trajectory translates to higher success.

7.1 Exploratory Temporal Analysis

First, we explore engagement trajectories with core course content over the duration of the course for both the self-paced and instructor-paced course.

The figure 7-1 visualizes the distribution of students engagement with lecture videos over the duration of the course in both the self-paced and Instructor-paced forum. The learners have been further broken down into those who completed the course and those who did not. The x-axis captures the number of unique videos watched in a specific week. The concept of the week does not have the same importance in self-paced course, nevertheless, it does represent a time unit that is narrow yet broad enough to capture engagement trajectory. The first thing that we notice here is an apparent difference between completers and dropouts. Dropouts are relatively engaged in the early few weeks of the course, but then their engagement falls drastically. The fall is more sudden in a self-paced course than in an Instructor-paced one. The other thing to notice from the figure is the difference between engagement trajectory of the completers across the different formats. The viewing pattern of completers in Instructor-paced course follows a normal distribution in most weeks except in the weeks of exams (6th and 12th). Interestingly, that is not the case with a self-paced course where the viewing pattern in most weeks looks like an exponentially decaying distribution. The tail of the distribution gets longer over time, suggesting that learners in the self-paced format ramp-up activities at the end of the course.

Next, we plot a similar figure 7-2 for problem submission over time . We focus here on the ungraded problems as the only graded assessment in self-paced version is the final exam. Overall, we see a similar trend here with some subtle differences. First, we notice that in an instructor-paced format, the drop in engagement is even more sudden than videos. Additionally, in the self-paced version, dropouts hardly engage with ungraded problems at any stage of the course. Next, we notice a very interesting shift in the way learners approach problems in the different course format. In an instructor-driven format, we see high engagement with the problems during the first half of the course, but then it falls in the next half. This behavior is opposite in the self-paced course, where we see an increase in engagement in the latter half of the course. If one hypothesizes that learners attempt ungraded problem to prepare for the graded assessments, then the apparent shift in self-paced assignment seems logical.

We conclude this discussion by summarizing the engagement trajectory of learners in instructor-paced and students-paced versions of SC0x in the figure 7-3 and 7-4. We show week relative to course start in the x-axis and plot the average number of videos watched per week on the y-axis. We also plot problems submitted and forum events on the same graph by standardizing these activities relative to videos watched. Completers behave very different across both the versions with completers in instructor-paced following a cyclic pattern of high early engagement, a dip around week 6 and resurgence in the latter part of the course.

On the contrary, completers in self-paced start slow but ramp up their activities at the end of the course. Dropouts in both the versions mostly follow the same pattern with dropouts in self-paced engaging less than their counterparts in instructor-paced version.

7.2 Clustering Method

Next, we formally describe our clustering method to determine longitudinal engagement trajectories in MOOCs. We discuss the method in this section and present the results in the subsequent sections.

Clustering for time series data is a well studied, yet a contentious topic [43, 32]. The concept of similarity distances does not translate to data varying across time. Caution, therefore, must be practiced to apply traditional clustering techniques on time series data. In MOOCs, several studies have explored clustering methods on clickstream data [36, 23]. The general approach has been first to compute a description for each student of the way in which the student has "engaged" throughout the duration of a course and then apply clustering techniques to find sub-populations within these engagement descriptions. One apparent drawback of this approach is that it requires one to manually describe the engagement style, which can bias the clustering method.

We instead rely on a data-driven approach and let our classification method determines the sub-population on its own. This fits with our problem statement well as we are interested to understand the difference in engagement trajectories across the different formats rather than to fit sub population into some user-defined groups.

We formulate the problem as sequence data where each X is a multivariate sequence with T samples. We then included contextual data associated with each sequence into the final formulation [17, 16].

We then proceeded as follows [16, 17]:

1. A sliding window was applied to the sequence data to transform it into a piecewise representation (segments). (width=2, overlap=0.5) [9]

2. For each segment, we computed the following features: mean, variance, standard deviation, maximum, minimum, skewness, kurtosis, mean crossings, mean spectral energy, and a 4-bin histogram.

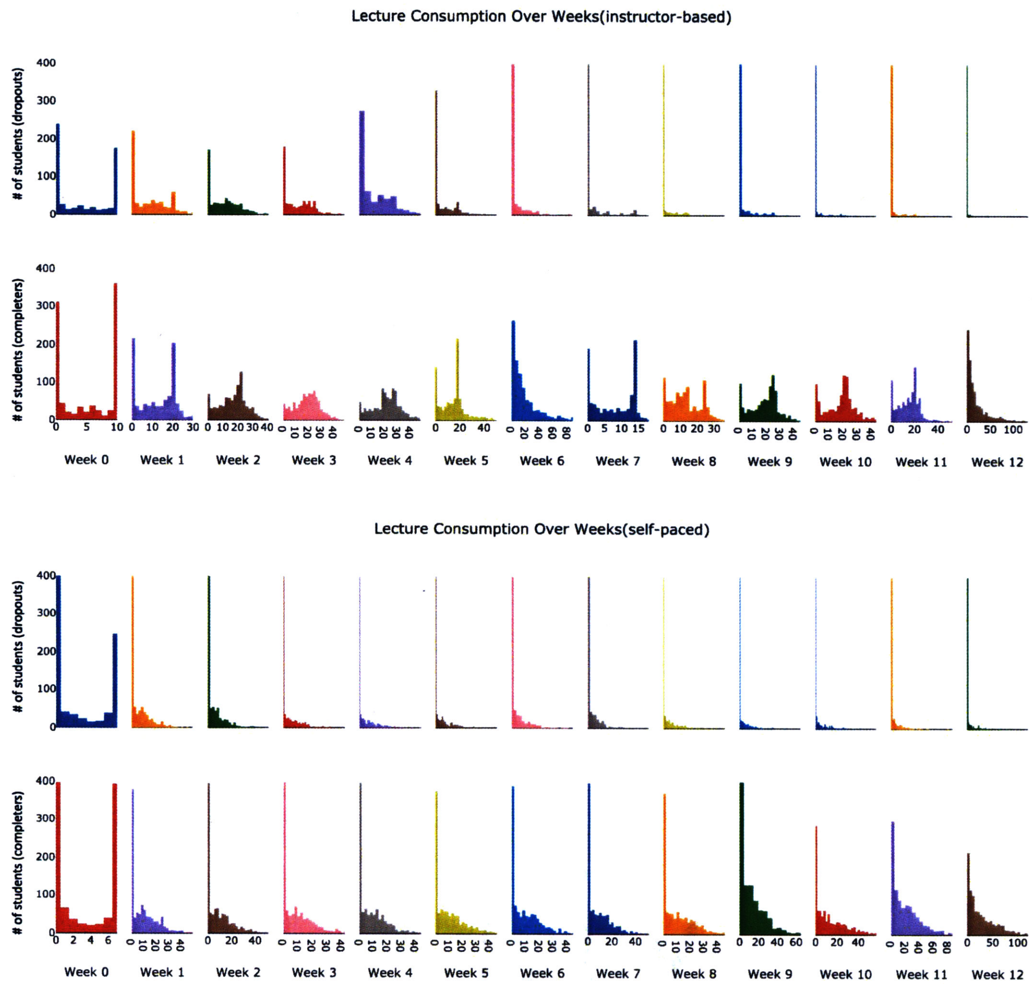


Figure 7-1: Weekly distribution of number of unique videos watched for completers and dropouts in instructor-paced (top) and self-paced version (bottom) of SC0x. We plot a distribution for each of the week where x-axis is the number of videos watched in that week and y-axis is the number of students.

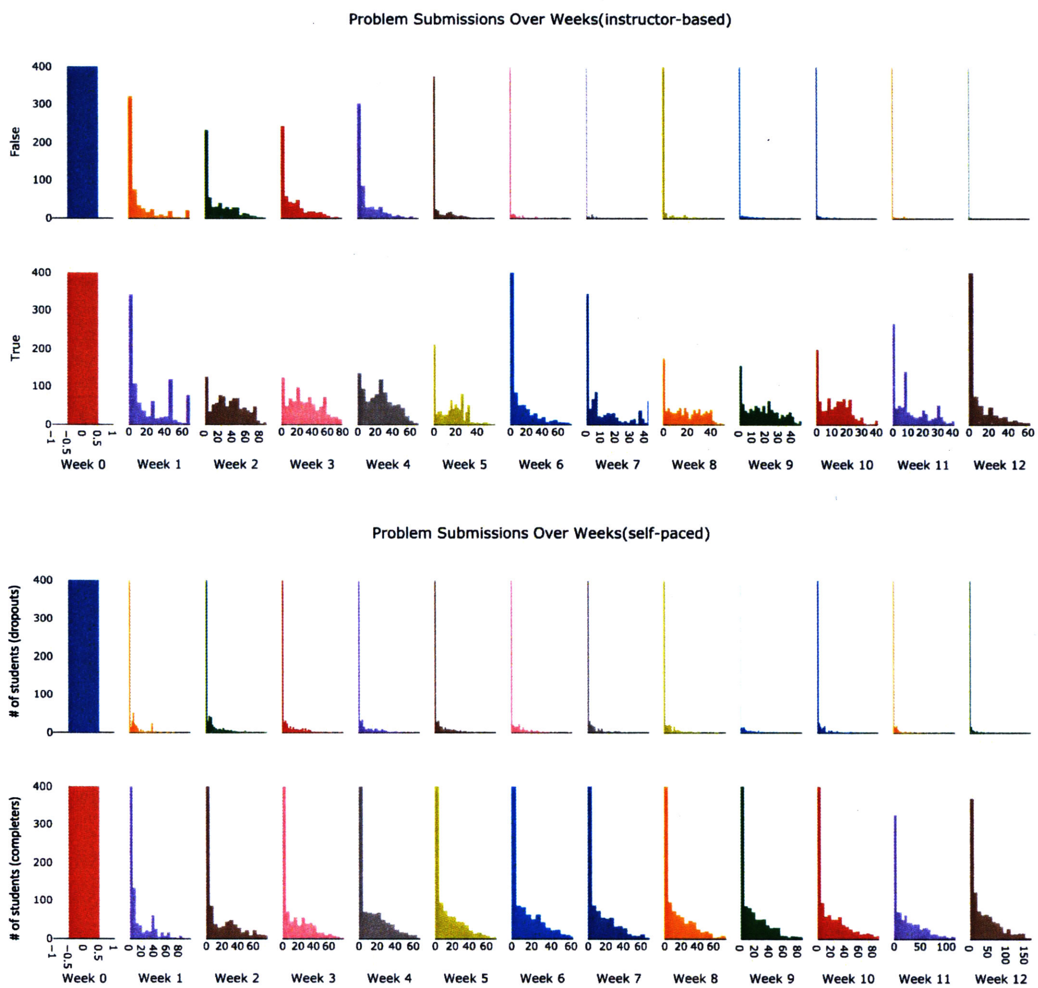


Figure 7-2: Weekly distribution of number of unique problems submitted for completers and dropouts in instructor-paced (top) and self-paced version (bottom) of SC0x. We plot a distribution for each of the week where x-axis is the number of problems submitted in that week and y-axis is the number of students.

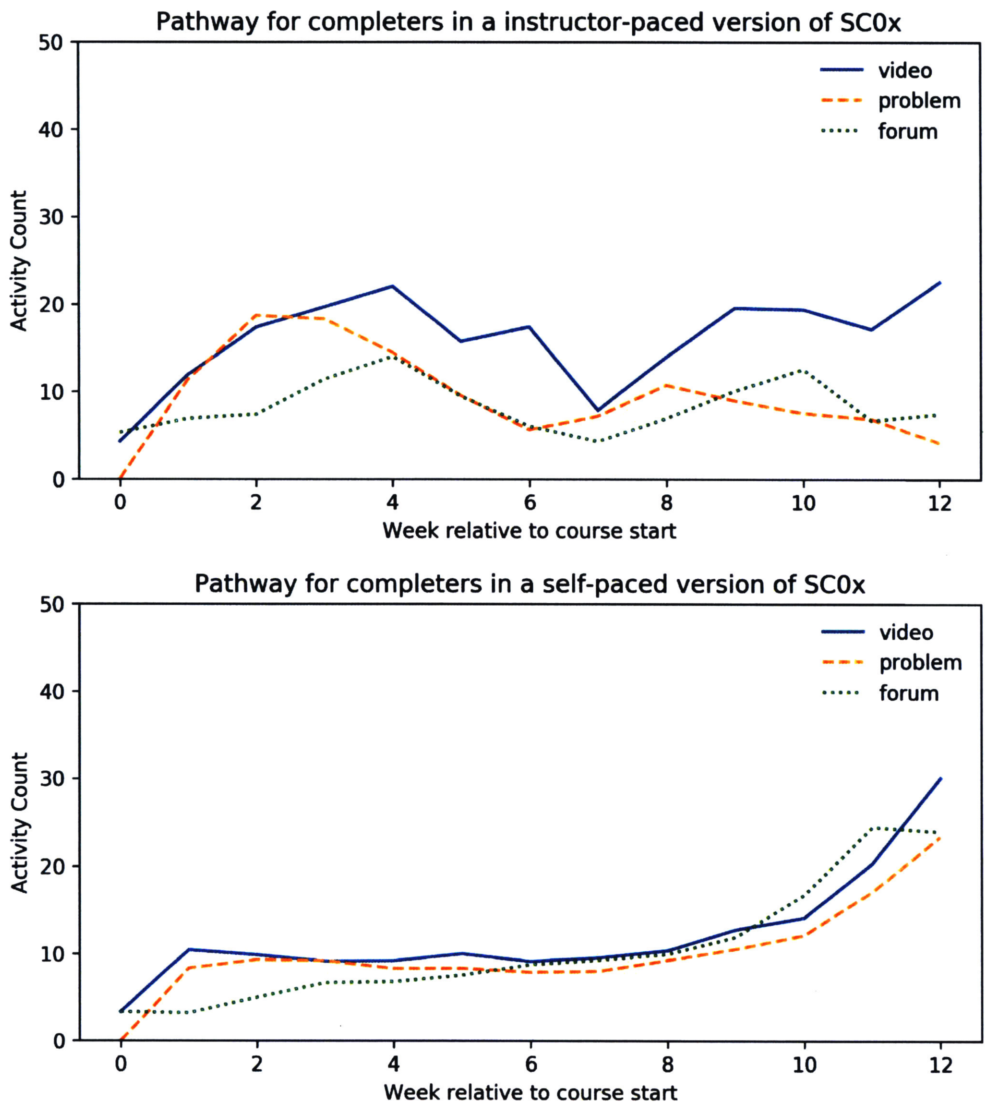


Figure 7-3: Engagement trajectory of completers in instructor-paced (top) and self-paced (bottom) versions of SC0x. X-axis shows week relative to course start and Y-axis shows average number of videos watched per week for completers and dropouts. We also plot problems submitted and forum events on the same graph by standardizing these activities relative to videos watched.

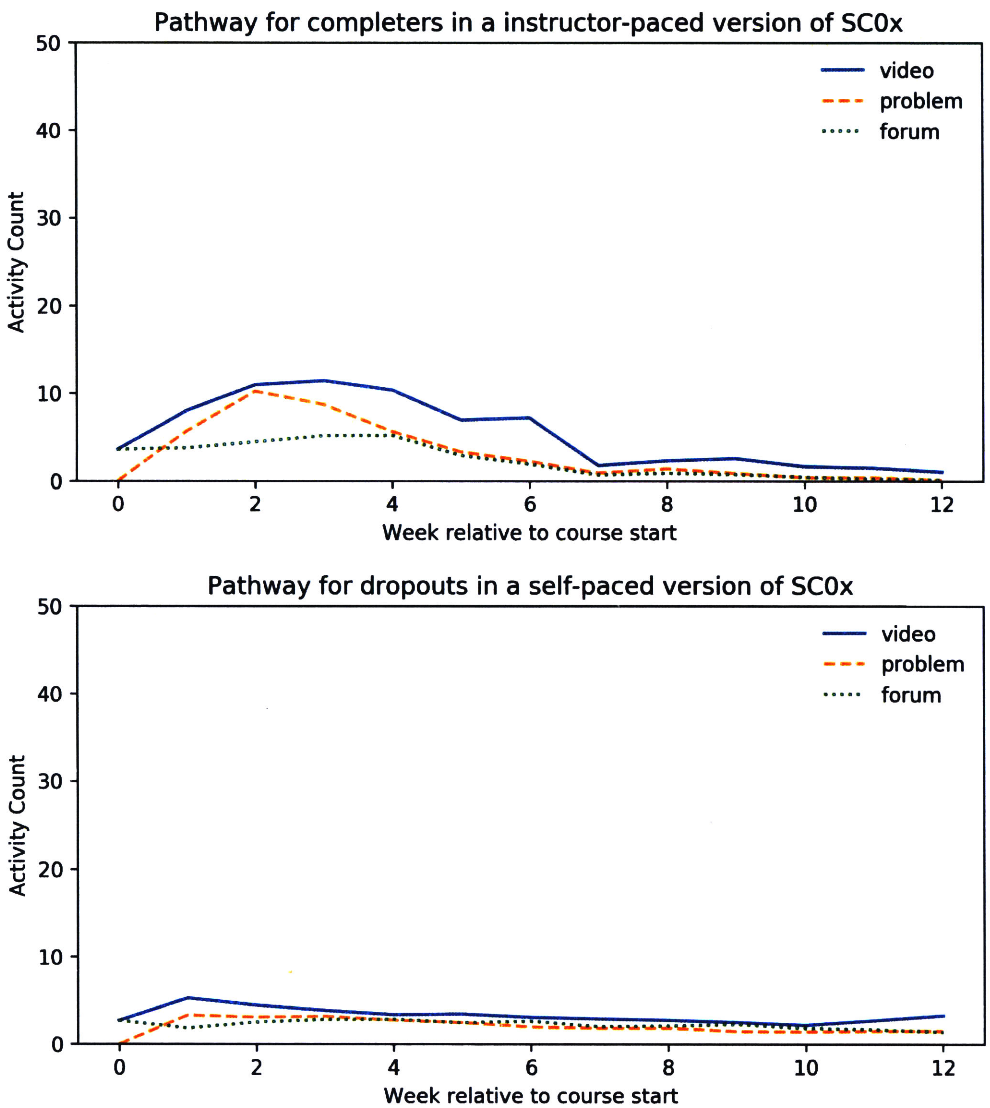


Figure 7-4: Engagement trajectory of dropouts in instructor-paced (top) and self-paced (bottom) versions of SC0x. X-axis shows week relative to course start and Y-axis shows average number of videos watched per week for completers and dropouts. We also plot problems submitted and forum events on the same graph by standardizing these activities relative to videos watched.

3. Features were scaled to zero mean and norm standard deviation, and a Gini importance ranking, feature selection strategy, was implemented to reduce the feature space [44].

4. A standard K-means clustering algorithm was applied to identify engagement pattern. We used heuristics and silhouette analysis to determine the number of K. K=4 gave us the cluster with different trajectories and enough resolution.

5. To account for the randomness of the K-means algorithm, the algorithm was repeated a hundred times, and the solution with the highest likelihood was selected.

The figure 7-5 summarizes the keys steps of the clustering method.

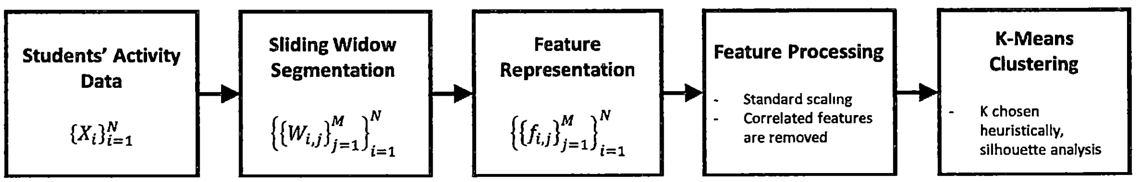


Figure 7-5: Summary of clustering method used to determine longitudinal engagement trajectories in SC0x. X_i : time series, N: number of time series in the data set, $W_{i,j}$: sliding window segment (derived from X_i), $f_{i,j}$: feature vector (calculated from $W_{i,j}$) [16, 17]

7.3 Engagement Trajectory for Completers

We now turn our attention to subpopulations identified by the clustering algorithm. First, we compare learners who completed the course across both the formats.

7.3.1 Completers in Instructor-paced course

The figure 7-6 displays the engagement trajectories of completers for the four sub-populations recognized by the classification algorithm in an instructor-paced course.

- **Instructor-Completers-Cluster1:** Cluster 1 is the least engaged among all the other clusters. While they engage relatively less than others, they show different engagement behavior with the various course contents. They do watch more than 70%

of the videos but engage very little with the problems (25% of ungraded problems) and forum content. We see a dip in problem engagement after the midterm. Perhaps they do not find assessments useful, or their excitement in the subject dies down with time.

- **Instructor-Completers-Cluster 2:** Cluster 2 has the most representation among the completers. They watch 96% of the videos and completes 65% of the ungraded problems. Their engagement with forums is, however limited. Their engagement behavior generally remains uniform during the duration of the course, excluding a dip in the problem activity at the end of the course, which is a general trend with the instructor-paced course. Overall, these learners secure better grade than cluster 0.

- **Instructor-Completers-Cluster 3:** Learners in Cluster 3 are the most engaged. They watch 98% of the videos and completes 76% of the ungraded problems. However, what distinguishes them from others is their extra-ordinary forum activities. They engage both actively and passively with the forum. They represent only 2% of the total learners who complete the course.

- **Instructor-Completers-Cluster 4:** Learners in Cluster 3 are all-rounders. They demonstrate relatively uniform engagement with all the course content. Their behavior also remains the same during the duration of the course. They represent 23% of the total learners who complete the course.

7.3.2 Completers in self-paced course

The figure 7-7 displays the engagement trajectories of completers for the four sub-populations recognized by the classification algorithm in a self-paced course.

- **Student-Completers-Cluster 1:** Learners in cluster 1 engage with the course in a cyclic manner. They demonstrate high early engagement but then midway, their activity slows down, which then jumps up again at the end of the course. In many sense, their behavior is typical of what was seen in an instructor-paced course-high early engagement, dip around week 6 and then a jump at the end of the course. Overall, they watch 88% of the videos and attempts 66% of the ungraded problems.

- **Student-Completers-Cluster 2:** Learners in Cluster 2 exhibit uniformity in engagement during the duration of the course. They follow a strict regiment of viewing videos and attempting problem every week. They also engage wholeheartedly with the forum demonstrating high social engagement. In total, they watch 94% of the videos and attempts 88% of the ungraded problem-highest among other clusters. Overall, they secure the highest grade among learners who complete the self-paced course.

- **Student-Completers-Cluster 3:** Learners in Cluster 3 engage with course content in an increasingly linear manner, i.e., they start slowly but then ramp-up their engagement linearly till the end of course. Engagement with all course content follows this trajectory. Ultimately, they end up watching 93% of the videos and attempting 66% of the ungraded problems.

- **Student-Completers-Cluster 4:** Learners in Cluster 4 initiate activities very late into the course (at around eight weeks) and then exponentially ramp up engagement in the last two weeks. Because they start late, they focus more on videos (watches 76% of the videos) and less on problems (26% of ungraded problems). The low completion rate for problems is reflected in their grade reports, and they secure the lowest grade among others.

The trajectories identified from instructor-paced course varies significantly from a self-paced course. Within an instructor-paced course, we found all sub-populations to follow the same trend- high engagement at the start, a dip around week 6 and then a slight resurgence at the end. There are some subtle differences between the sub-populations, but more or less, all of them demonstrate the same overall behavior. This is in contrast with the self-paced course, where each of the sub-populations takes on a diverse viewpoint towards the course. Cluster 1 demonstrates a cyclic behavior; cluster 2 engage uniformly with the content; while the latter two clusters start late and ramp-up activities linearly and exponentially till the end of course. The instructor-paced course had somewhat regulated the pathway of learners in the course. The shift to the self-paced course had evoked the natural or preferred learning pathway of the learners.

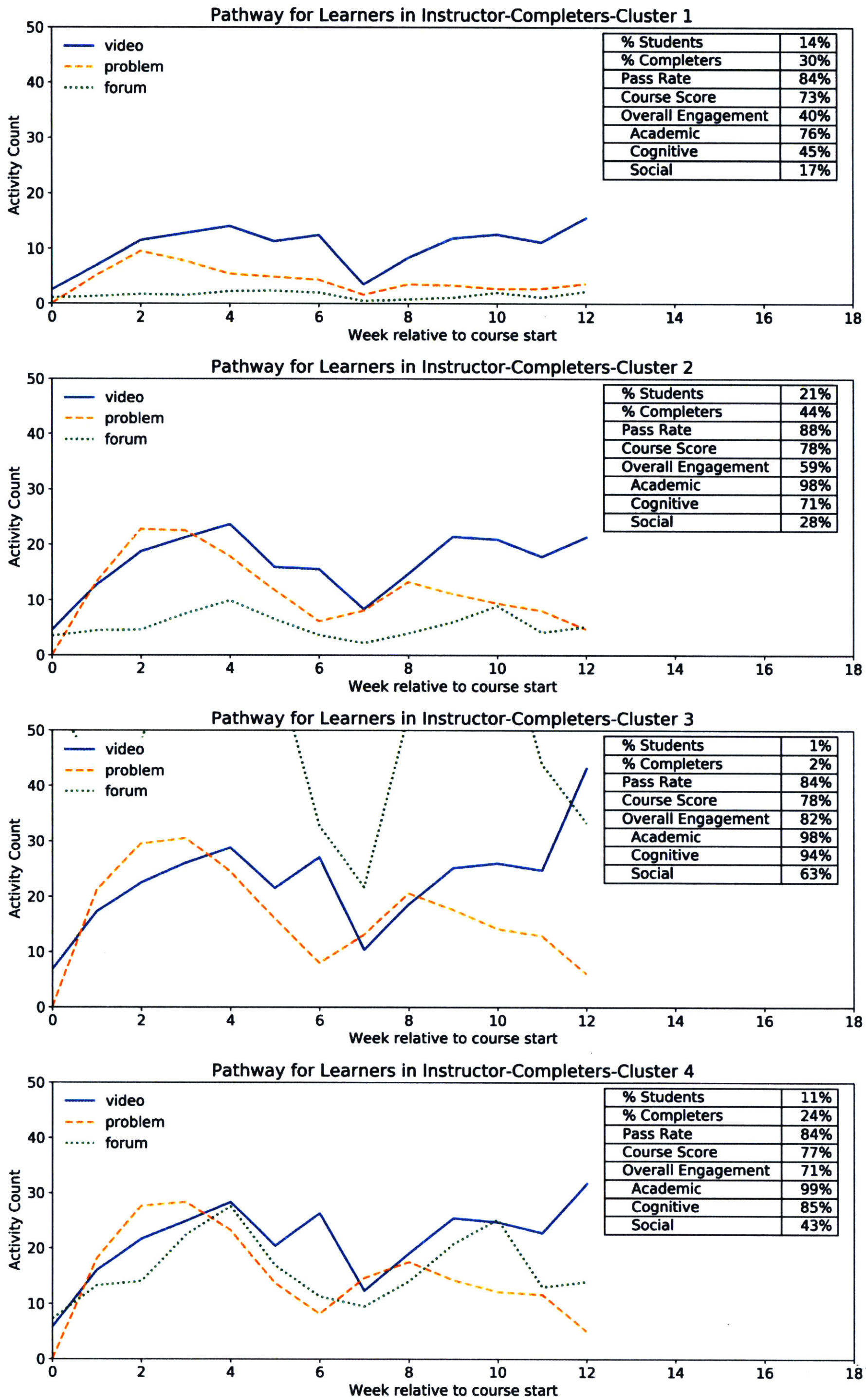


Figure 7-6: Clustering engagement trajectories for completers in instructor-paced versions of SC0x. Each graph represents a cluster with X-axis showing week relative to course start and Y-axis showing average number of videos watched per week for learners in that cluster. We also plot problems submitted and forum events on the same graph by standardizing these activities relative to videos watched.

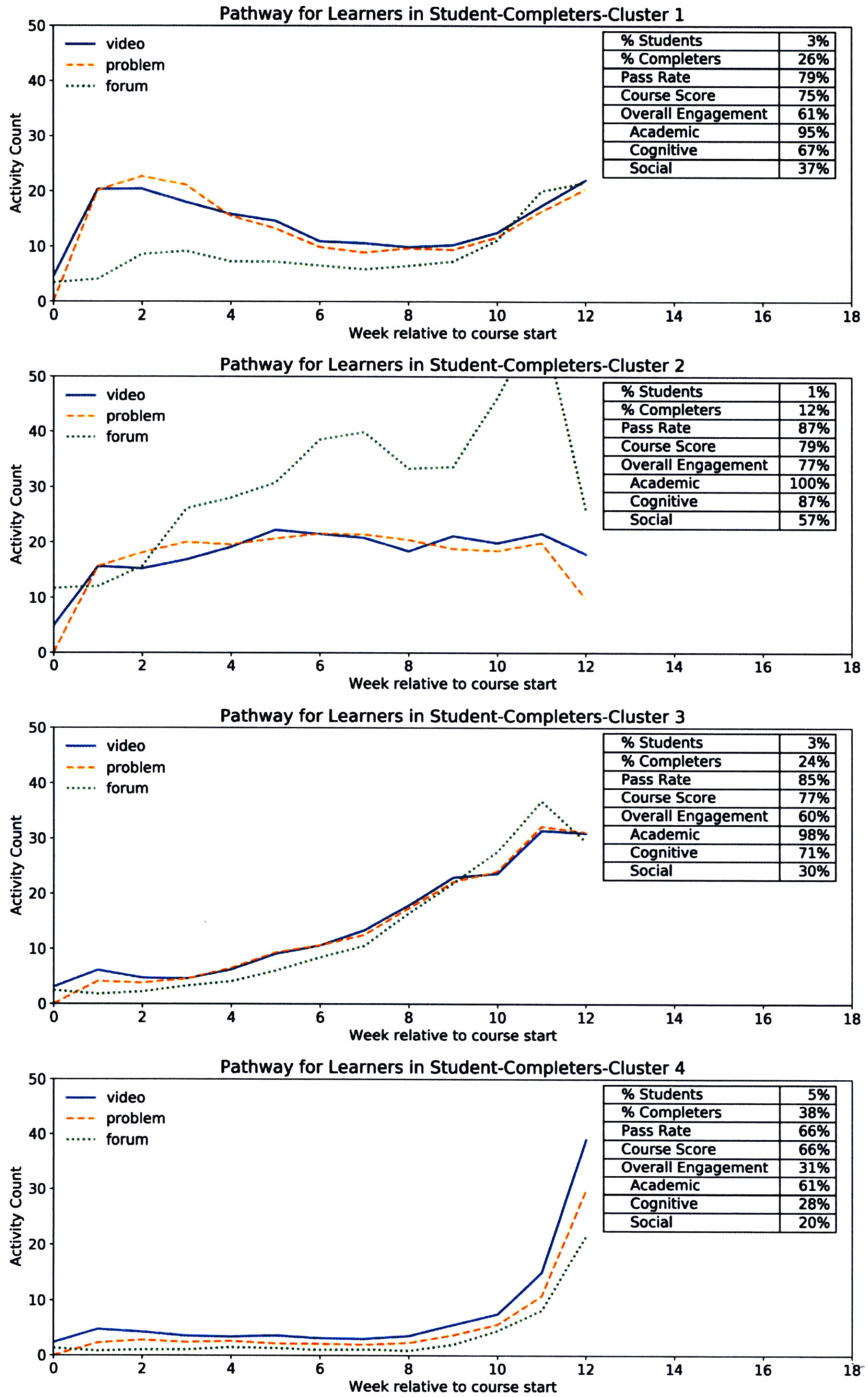


Figure 7-7: Clustering engagement trajectories for completers in self-paced versions of SC0x. Each graph represents a cluster with X-axis showing week relative to course start and Y-axis showing average number of videos watched per week for learners in that cluster. We also plot problems submitted and forum events on the same graph by standardizing these activities relative to videos watched.

7.4 Engagement Trajectory for Dropouts

Having analyzed clusters of completers, we turn our attention to dropouts. Their engagement style differs substantially from completers, and that is why it makes sense to examine them separately.

7.4.1 Dropouts in Instructor-paced course

We start by looking at dropouts in the instructor-paced course.

- **Instructor-Dropouts-Cluster 1:** They start well but then disengages quickly with many learners dropping out before or right after the midterm. 30% of these learners complete the midterm securing around 40% grade in the midterm exam.

- **Instructor-Dropouts-Cluster 2:** These learners rarely engage in the course. Only 6% of them completes the midterm. Looking at their trajectory, one can say that these learners make their decisions to quit early in the course (first/second week).

- **Instructor-Dropouts-Cluster 3:** They exhibit high engagement with the course content in the first half of the course, but then their activity falls significantly after the midterm. A characteristic trait of these learners is their high participation with the forums. 93% of them gives the midterm exam, and on average they secure a 54% score in it. Perhaps, the lower grade in the midterm discourages these learners from completing the course.

- **Instructor-Dropouts-Cluster 4:** These learners are similar to cluster2, but they display less engagement with forums. Their activities in the first half of the course are high, but it dips down in the second half. 88% of them participates in the midterm exam and on average they secure 47% score in it, suggesting that the lower grade in the midterm discourage them from completing the course.

7.4.2 Dropouts in self-paced course

We now conduct a similar exercise for dropouts in self-paced course:

- **Student-Dropouts-Cluster 1:** They engage moderately with videos but very little with problems and forums. Also, while they watch videos regularly, their overall activity level remains moderate during the duration of the course. They represent 23% of the total dropouts.

- **Student-Dropouts-Cluster 2:** They demonstrate hardly any engagement with the course. They represent 47% of the total dropouts.
- **Student-Dropouts-Cluster 3:** They exhibit moderate to high engagement in the first few weeks of the course but then disengages, ultimately dropping down to zero activity by the mid of the week. They represent 22% of the total dropouts.
- **Student-Dropouts-Cluster 4:** They display consistent engagement during the duration of the course. They remain engaged until the end of the course but decide not to take the final exam. Overall they watch 60% of the videos and attempts 40% of the problems. They represent 8% of the total dropouts.

The dropout behavior between the different course format shows both similarity and differences. First, we make a point that the dropouts in self-paced courses engage less than those in the instructor-paced course. There is a large body of dropouts that rarely engage in a self-paced course. Dropouts in instructor-paced version do show some engagement even if it is little or moderate. Secondly, we see some common trends in both types of courses: moderate to high early engagement at the start and a fall in the engagement by the mid of the course. Thirdly, in both cases, we do see dropouts that maintain activity throughout the course.

7.5 Clusters and Demographics

Now that we have looked at the engagement trajectories, we want to explore if these trajectories represent learners with certain demographics features or not. In other words, we are interested to know if certain users are inclined towards a particular trajectory or not.

7.5.1 Method

To answer this question, we build a multinomial logistic regression with "cluster type" as the output variable and demographic features as predictors. The features were discretized, and highly correlated features were removed. We experimented with two models, one that

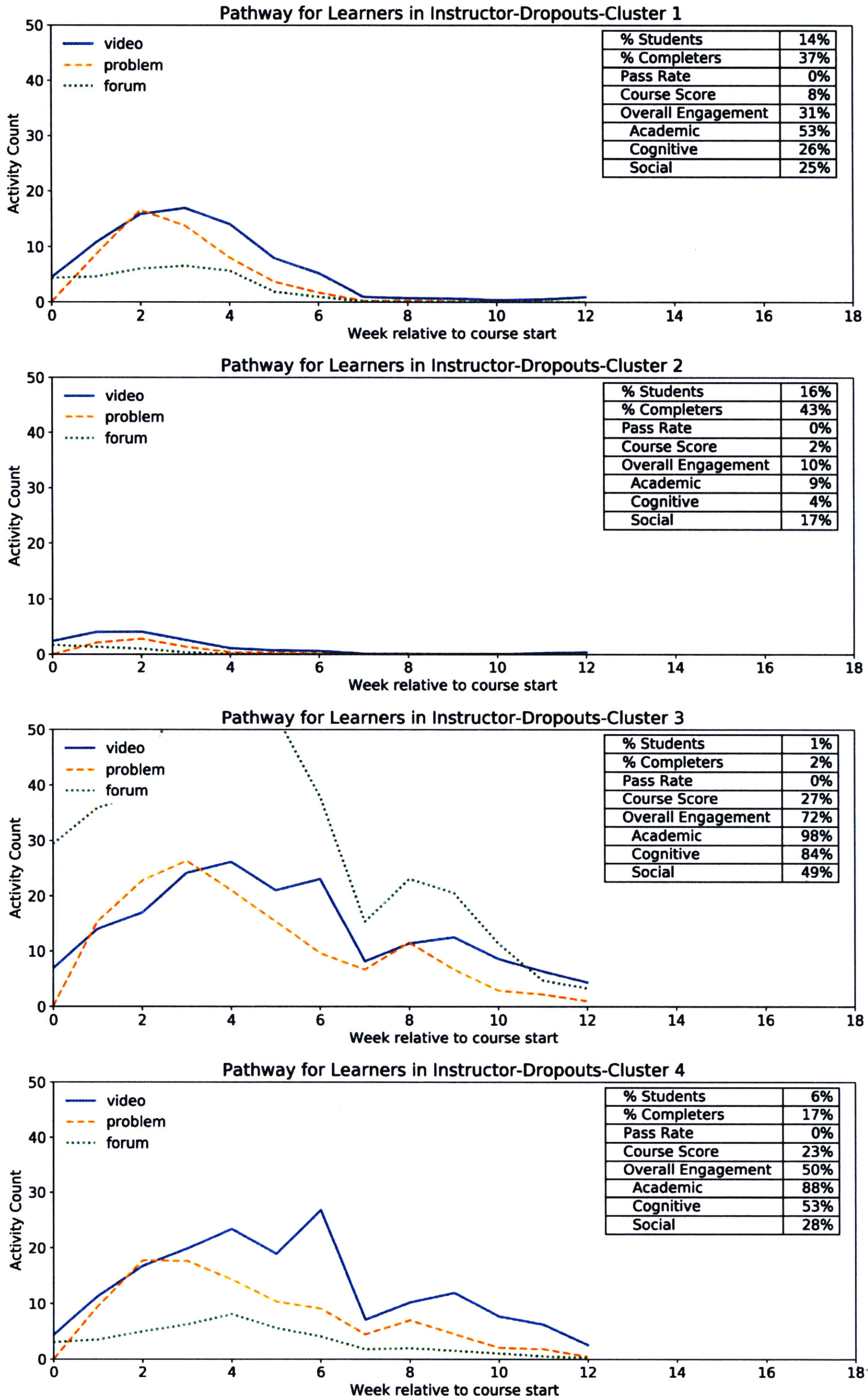


Figure 7-8: Clustering engagement trajectories for dropouts in instructor-paced versions of SC0x. Each graph represents a cluster with X-axis showing week relative to course start and Y-axis showing average number of videos watched per week for learners in that cluster. We also plot problems submitted and forum events on the same graph by standardizing these activities relative to videos watched.

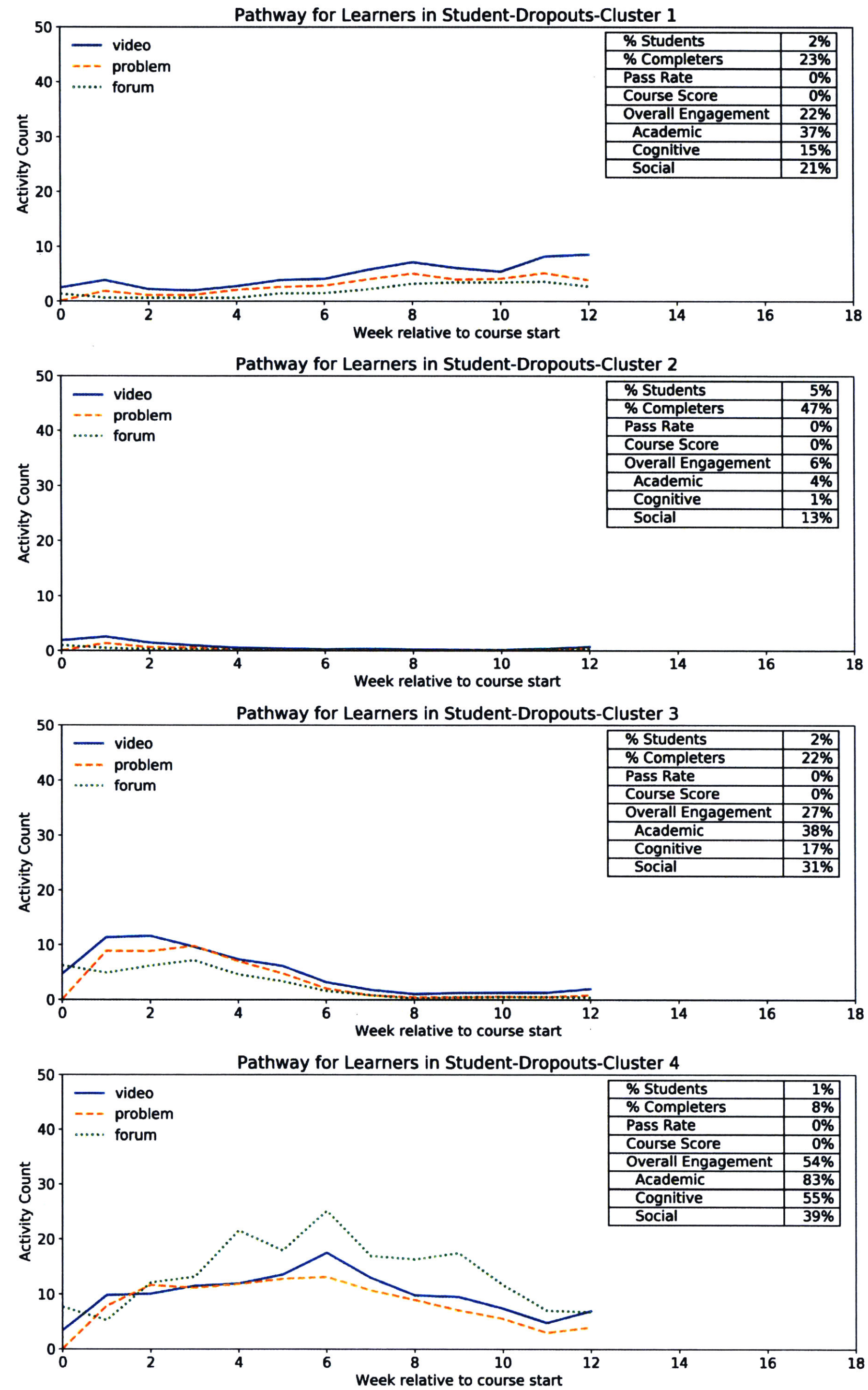


Figure 7-9: Clustering engagement trajectories for dropouts in self-paced version of SC0x. Each graph represents a cluster with X-axis showing week relative to course start and Y-axis showing average number of videos watched per week for learners in that cluster. We also plot problems submitted and forum events on the same graph by standardizing these activities relative to videos watched.

only included demographic features from the platform and the other that included features from the entrance survey as well. We applied this model to both completers and dropouts in the self-paced format of SC0x.

7.5.2 Result

The table 7.1 and 7.2 shows the output of the logistic regression. We convert the relative probabilities to marginal effects to better interpret the results. We define a reference learner group and compare clusters with the reference learner. Our reference learner has the following characteristics: gender (male), education (bachelor), age (30-40), english_index (70<x<90), hdi_index (75<x<90), verification date (first week of course), empstatus (employed), familiarity (Somewhat familiar), fluency (fluent).

first share the results of clusters identified within completers.

- **Student-Completers-Cluster 1:** There is 6% more probability for a female to be in cluster 1 than male. Also, we notice that learners that complete course verification earlier has 25% higher odds to be in this cluster than learners who complete verification in the first week of course. Despite showing strong intent, these learners decide not to participate in the exams. Perhaps, they struggle with managing their time during the duration of the course.

- **Student-Completers-Cluster 2:** Learners in these clusters have a demographic profile opposite to cluster 0. Females have 9% less probability of being in this cluster, and learners that complete verification at the latest date has 30% odds to follow this trajectory. Like cluster 1, it seems that these learners struggle to find time to study for the course.

- **Student-Completers-Cluster 3:** Learners with more familiarly with the topic have 60% higher probability of being in this cluster than our reference learner.

- **Student-Completers-Cluster 4:** Learners who are slightly older (40+) have 7% more probability of being in this cluster.

We now share the results of clusters identified within dropouts.

- **Student-Dropouts-Cluster 1:** There is 6% more probability for a female to be in cluster 1 than male. Learners in the age bracket between 20 and 30 are 6% less likely to be in this cluster than learners in the age bracket between 30 and 40. Also, we notice that learners that complete course verification before the start of the course have 10% higher odds to be in this cluster than learners who complete verification in the first week of course. In summary, professional females with a strong intent to complete the course are more likely to chose this trajectory than our reference learners.

- **Student-Dropouts-Cluster 2** Learners who are slightly older (40+) have 8% more probability of being in this cluster. Unemployed learners are 10% more likely to follow this trajectory than those that are employed. Also, learners that are very familiar with the content have 22% fewer odds to be in this cluster. Overall older, unemployed, and moderately familiar learners chose this pathway.

- **Student-Dropouts-Cluster 3:** Learners with higher math skills and more familiarly with the topic have a higher probability (8% and 12% respectively) to be in this cluster than our reference learner. These learners also show strong intent by completing the verification process within the first week of the course.

- **Student-Dropouts-Cluster 4:** Learners who are slightly younger (20-30) have 7% more probability of being in this cluster. College Students are 14% more likely to follow this trajectory than those that are employed. Also, learners in this cluster are more likely to complete the verification later (19%) in the course than doing it in the first week of the course.

7.6 Clusters and Outcome

Lastly, we examine the effect of engagement trajectory on course outcomes to examine how these different pathways relate to student course performance.

7.6.1 Method

To begin to address this question, we estimated a logistic regression model examining the relationship between trajectory clusters and course score. We discretized all the features and defined a reference learner for interpretation. Our reference learners has the following characteristics: gender (male), education (bachelor), age (30-40), english_index (70<x<90), hdi_index (75<x<90), verification(first week of course), empstatus (employed), familiarity (Somewhat familiar), fluency (fluent)

7.6.2 Result

The table 7.3 shows the output of the logistic regression. First, we present a baseline model that includes user demographics and course registration features. We find that there is a significant, positive relationship between the level of education, countries of residence and age. Learners that do not have college degrees have 20% less probability of passing the course than learners in the reference group. Learners in the age bracket between 20 and 30 also have a slight edge of passing the course than the reference group. Finally, we notice that learners from countries with good math skills have 16% more probability of passing the course.

In the next model, we include an additional term for the cluster type. We found that all cluster types are statistically significant and have a large impact on the pass rate. Cluster 0 (completers) is chosen as the reference. Learners who follow cluster 1 have 22% more odds of passing the course than learners who follow the pathway in cluster 0. Similarly, Learners in Cluster 1 have 14% probability of passing than our reference learners, suggesting that efforts spent at the end are more positively correlated with performance in the exam. On the flip side, learners in cluster 3 have 15% less probability of passing the exam than learners in cluster 0.

7.7 Conclusion

In this chapter, we explored the impact of pacing on learners' pathway during the duration of the course. The shift to self-paced delivery caused students to behave differently than those that have undertaken the course with instructor pacing. In an instructor-paced version, most students proceeded in a similar manner with high engagement at the start, a dip around midterm exam and an upturn in the second half of the course. In a self-paced version, students pursued their own style of trajectory through the course. Some took a cyclical pathway, while others remained uniformly engaged with the material. There were others that

MNLogit Marginal Effects

cluster=0	Covariates	dy/dx	P>|z|
Model 1	age: 20-30	-7%	0.042*
Model 1	verified: 30 days before course	-11%	0.005**
cluster=1	Covariates	dy/dx	P>|z|
Model 1	age: 40-50	8%	0.006**
Model 1	hdi_index: 0.5-0.7	-5%	0.081
Model 1	verified: 14 days before course	6%	0.083
Model 2	familiarity: Extremely familiar	-25%	0.054
Model 2	empstatus: Unemployed	10%	0.014*
Model 2	fluency: Proficient	5%	0.09
cluster=2	Covariates	dy/dx	P>|z|
Model 1	english_index: 75-100	-9%	0.037*
Model 1	verified: 14 days before course	-23%	0.002***
Model 2	familiarity: Extremely familiar	14%	0.065
cluster=3	Covariates	dy/dx	P>|z|
Model 1	age: 20-30	7%	0.031*
Model 1	verified: after 30 days of start of course	19%	0***
Model 2	empstatus: Full-time student	14%	0.03*

Table 7.1: Output of logistic regression models predicting cluster type using demographic features for completers in self-paced version of SC0x. Model 1 contains demographic data and Model 2 contains survey data while controlling for the demographic data. Truncating covariates whose p-value is greater than 0.1.

started slow and linearly increased their engagement while others started near the end of the course and ramped up their activity exponentially. These contrasting trajectories impacted students' grade, with "uniform" and "linear increase" styles having a positive correlation with course grade. Learners who took the "exponential" style performed the worst. Learners with certain characteristics were inclined to follow a specific pathway with older learners pursuing a "uniform" style and college-going students taking on an "exponential" type of engagement.

MNLogit Marginal Effects

cluster=0	Covariates	dy/dx	P>|z|
Model 1	english_index: >90	-11%	0.008**
Model 1	hdi_index: 50-75	8%	0.037*
Model 1	verified: 30 days before course start	25%	0***
Model 2	familiarity: Very familiar	-9%	0.065
cluster=1	Covariates	dy/dx	P>|z|
Model 1	verified: 30 days after course start	30%	0***
Model 2	familiarity: Very familiar	11%	0.044*
Model 2	fluency: Intermediate	-27%	0.014*
cluster=2	Covariates	dy/dx	P>|z|
Model 1	english_index: 50-75	-10%	0.066
Model 2	verified: 30 days before course start	-20%	0***
cluster=3	Covariates	dy/dx	P>|z|
Model 1	edu: hs	5%	0.092
Model 1	age: 40-50	6%	0.02*
Model 2	fluency: Intermediate	9%	0.035*

Table 7.2: Output of logistic regression models predicting cluster type using demographic features for completers in self-paced version of SC0x. Model 1 contains demographic data and Model 2 contains survey data while controlling for the demographic data. Truncating covariates whose p-value is greater than 0.1.

Logit Marginal Effects

Dep. Variable: y
Method: dydx
At: overall

	Model 1 dy/dx	Model 1 P>|z|	Model 2 dy/dx	Model 2 P>|z|
gender: f	2%	0.468	1%	0.772
edu: a	-15%	0.089	-13%	0.109
edu: hs	-19%	0.008**	-17%	0.01*
edu: m	7%	0.014*	6%	0.016*
edu: p	-16%	0.159	-12%	0.314
age: 0-20	32%	0.29	-6%	0.805
age: 20-30	8%	0.008**	7%	0.009**
age: 40-50	2%	0.722	-2%	0.654
age: 50+	8%	0.307	0%	0.971
english_index: 50-75	-13%	0***	-11%	0.001***
english_index: 90+	-22%	0***	-18%	0***
hdi_index: 0.5-0.7	-15%	0***	-16%	0***
hdi_index: 0.9-1	6%	0.193	6%	0.19
verified: 30 days before	17%	0***	14%	0.001***
verified: 14 days before	14%	0.022*	12%	0.034*
verified: 14 days after	3%	0.496	2%	0.658
verified: within 30 days	8%	0.246	4%	0.525
verified: after 30 days	4%	0.507	5%	0.304
enrollment: 30 days before	-14%	0.017*	-10%	0.089
enrollment: 14 days before	-10%	0.128	-10%	0.132
enrollment: 14 days after	1%	0.853	6%	0.183
enrollment: within 30 days	2%	0.726	4%	0.353
enrollment: after 30 days	3%	0.516	6%	0.13
cluster_1			22%	0***
cluster_2			14%	0***
cluster_3			-15%	0***

Table 7.3: Output of logistic regression models predicting course score using trajectory types for completers in self-paced version of SC0x. Model 1 contains demographic data and Model 2 contains trajectory types while controlling for the demographic data. Truncating covariates whose p-value is greater than 0.1.

Chapter 8

Proficiency Measurement: Impact of course pacing on student success

In this section, we investigate the impact of course delivery mode on student success. We define student success with three core metrics: course completion, course certificate, and course engagement. For a detailed discussion on these metrics, we refer readers to chapter on our methodology. Having defined these metrics, we are interested in evaluating the effect of course delivery on these student metrics.

Our work up till now has established an association of specific variables with course delivery modes. For example, in the previous section, we showed that the self-paced course is associated with lower engagement. We have, however, shied away from claiming that self-paced course can cause lower student success or vice versa. We can only make such a claim if we had established a causal relationship between them.

The cleanest way to measure the effect of course delivery mode is by running a randomized control trial. Specifically, we want to randomize who gets enrolled in a self-paced course and who in the instructor-paced course, and then look at the student success metrics. This removes the effect of any confounding variables which might be influencing the student success metrics. In ours and many other educational situations, running an A/B test to directly measure the effects of an intervention is impractical, unfeasible or unethical. Given our situation of not being able to do A/B test, how can we measure the impact of an intervention on student success?

To answer this question, we use a potential outcomes framework to try and make causal inferences about situations such as ours, where we only have observational data [63]. Fundamentally we are interested in knowing how a self-paced learner would have performed if it had participated in instructor-paced course. Since this is something we never physically observed, we will use statistical and econometric methods to answer this question.

8.1 Method

We begin with formally setting up the problem. To be consistent with the language used in the potential outcomes framework, we use the following terminology:

- Learners are called as "subject"
- A control group is one where the subjects (learners) participated in a MOOC with an instructor-paced format.
- A treatment group is one where the subjects (learners) participated in a MOOC with a self-paced format.
- Student success metrics are called as "potential outcomes".
- Average treatment effect is the mean difference in student success metrics of a set of individuals that have taken both self-paced and instructor paced course.

With this new language setup, we can now mathematically formulate the problem.

- Y represents the potential outcome where Y_0 is the potential outcome of a subject in the absence of treatment, and Y_1 is the potential outcome when treated.
- X is the treatment status with $X = 1$ indicating treatment and $X = 0$ indicating control.
- If we had been able to run A/B test, we would have randomly assigned students to either self-paced or instructor-paced format. This makes the outcome independent of the treatment i.e. $E[Y_1|X = 1] = E[Y_1]$. In our situation, no random assignment can be be made and thus this equation doesn't hold. We overcome this by using other information we have about the students. Let us call this information, Z (known as covariates). We then assume that Z includes enough additional information to explain completely the choice of a student to enroll in a self-paced course.

Potential Outcome	Completion, Pass rate, Engagement Index
Treatment	self-paced
Control	Instructor-paced
Covariates	Demographic: gender, age, education, employment, country, hdi, os Timing: verification data, enrollment data, edX joining date Skill: language proficiency, math skills, familiarity with course content Grade: Pass threshold on final exam score

Table 8.1: Summary of elements used in potential outcomes framework

We use covariates to model the potential outcomes for every student that has either taken the instructor-paced or self-paced course [5]. Specifically for every student,we estimate:

$$Y_0(Z) = E[Y|Z, X = 0] \tag{8.1}$$

$$Y_1(Z) = E[Y|Z, X = 1] \tag{8.2}$$

Once we have the potential outcomes, estimating the average treatment effect is simple. Mathematically:

$$ATE = 1/N \sum_i (Y_1(Z) - Y_0(Z)) \tag{8.3}$$

The list of covariates used for modeling is shown in table 8.1. Most features are self-explanatory, but two of them warrant discussion: math skill and grade. We use percentage correct on the first attempt in the early weeks as a proxy for math skills. Our hypothesis is that previous knowledge and prior skills play a significant role in the early success of the course than the engagement with the online course. Also, We compute a binary variable that indicates whether a student has passed the course by applying a threshold on the final exam score. We add this variable to compensate for any potential benefit that learners in instructor-paced course might gain from multiple graded assignments distributed throughout the course.

We estimate 8.1 using three methods. We refer readers to [30, 2, 28] for more details on these methods. We present a brief summary of each of the method here:

- **linear regression:** we use standard ordinary least square regression where coefficient of the treatment variable is computed as the causal effect.

- **Matching:** In Matching, we find for each student which has taken a self-paced course, a 'similar' student which has taken a instructor-paced course. Comparing the success metrics for these students then give us the average treatment effect. Nearest neighbour is used to estimate similarity.

- **Propensity score stratification:** we stratify the data into bins with identical common covariates and then compute the effect of treatment. We review this method in more detail in the next section.

8.2 Result

We estimated average treatment effect for each of the student success metrics. The tables 8.4 and 8.3 presents the causal effect of self-paced format on the learners. We computed these results using two different approaches. In the first approach, we only used data of the courses starting from 2018 so as to not to confuse the treatment under study with an earlier intervention in 2017. In the second approach, we included all of the data but formulated the problem to include the intervention in 2017 as a covariate. In both cases, the causal effect remained the same.

The result shows that self-paced format has caused degradation across all the student success metrics with pass rate experiencing the greatest decline of 10% as compared to instructor-paced format. Completion rate reduced by 6% and engagement index decreased by 8%. We also notice that the native estimate is not that far from the causal estimate, confirming that student demographics, skillset, and intent has remained mostly the same between these courses. Nevertheless, by controlling for these changes, even if they were minimal, we have reached to the more accurate impact of the course delivery mode on the learners.

Course Run	Total Enrolled	Total Verified	Verified Conversion	Course Completion	Pass Rate	Engagement
20171T	27,187	1,636	6%	50%	86%	53%
20173T	29,842	1,856	7%	55%	81%	57%
20181T	20,062	1,565	8%	65%	95%	59%
20183T	15,177	1,475	11%	61%	88%	59%
20191T	20,434	1,827	10%	55%	76%	51%

Table 8.2: Student success metrics for all runs of SC0x. Verified learners pay a fee to earn a certificate at the end. Completion rate is computed for verified learners. Pass rate is computed for verified learners that completes the course. Readers are referred to chapter 3 for more details on engagement metric.

	Completion		Pass Rate		Engagement	
	ATE	p-value	ATE	p-value	ATE	p-value
Naïve Estimate	-3%	0.023	-11.0%	0.0232	-6.1%	0.001
linear_regression_estimator	-7%	0.001	-10.8%	0.001	-8.1%	0.001
Propensity Score Matching	-7%	0.001	-10.9%	0.001	-7.9%	0.001
propensity_score_stratification	-5%	0.001	-10.4%	0.001	-6.4%	0.001
Causal Estimate	-5%		-10.4%		-7.5%	

Table 8.3: Impact of self-pacing on student success metrics. Naïve Estimate is calculated using all previous instructor-paced versions (2017-present). Completion rate only includes verified learners and Pass rate includes verified learners that have completed the course.

	Completion		Pass Rate		Engagement	
	ATE	p-value	ATE	p-value	ATE	p-value
Naïve Estimate	-9%	0.001	-15%	0.001	-8%	0.011
linear_regression_estimator	-7%	0.001	-11%	0.001	-8%	0.001
Propensity Score Matching	-7%	0.001	-9%	0.001	-9%	0.001
propensity_score_stratification	-7%	0.001	-10%	0.001	-8%	0.001
midrule Causal Estimate	-7%		-10%		-8%	

Table 8.4: Impact of self-pacing on student success metrics. Naïve Estimate is calculated using instructor-paced versions starting from 2018 (2018-present). Completion rate only includes verified learners and Pass rate includes verified learners that have completed the course.

8.3 Heterogeneous treatment effects

In this section, we investigate the impact on course delivery mode on various sub-populations within the course. Specifically, we are interested in knowing how similar groups within the course responds to course delivery modes. Notice the difference with the clustering approach we used in the temporal analysis where the focus was to identify a group of learners that follow a similar engagement trajectory. Here we are interested in classifying learners on their prior demographic and registration features with the goal to compare their performance across the two types of course delivery modes.

We cluster learners based on their demographic and registration features(covariates). Instead of using traditional clustering techniques, we rely here on propensity score matching methods. The propensity score is nothing but the probability that a learner will receive treatment. Rosenbaum and Rubin showed that for the subjects that share the same propensity score, the difference between the treated and the control units determines the average treatment effect [30, 73]. Thus instead of matching on the covariate vectors themselves, we can match on the single-dimensional propensity score, aggregate across subjects, and still arrive at a valid estimate of the overall average treatment effect.v

Once we have the propensity score, we use that score to stratify the learners into blocks with similar propensity scores. We use a data-driven procedure for selecting both the number of blocks and their boundaries, with the expectation that the number of blocks should increase with the sample size [30, 73].

The results for completion metric is given in table A.1 and A.2. The row shows the demographics features and the columns show the clusters identified using the above-mentioned method. We compare these clusters with the mean values of the overall group excluding that cluster; essentially we are trying to find what set these groups different from the others.

At a high level, we notice a couple of important observations. First, we see a large variation in completion rate across the clusters with the rate varying from -10% to 10%, indicating that a self-paced have impacted sub-populations differently. Secondly, there is a great deal of overlap between the clusters, showing the dropout is a complex phenomenon with many complex variables in play. Nevertheless, we do see that certain demographic features are prominent in a certain cluster, implying that learners with certain demographic features perform better in instructor-paced format and vice versa. To understand this impact, we focus on sub-populations that are on the extreme edges of the competition rate.

Learners in cluster 6 and 5 performed or would have performed better or almost the same in self-paced format than instructor-paced version of SC0x. There are more learners in these clusters that have the following attributes (when compared with the overall population):

- More familiarity with the course content.
- Increased experience with the platform and possibly with online courses.
- Comes from developing nations, which is reflected by lower English index, lower mac user and higher learners from developing countries(India, Egypt and others)
- Have bachelor degree
- Have specified their employment status as "unemployed."

Note that we don't claim that these attributes lead to higher success in self-paced students. We are just saying that if we had run an A/B experiment with learners from these groups, learners in self-paced format would have performed better or almost the same than instructor-paced version of SC0x.

Learners in cluster 0 and 1 performed the worst in self-paced format when compared with instructor-paced version of SC0x. There are more learners in these clusters that have the following attributes (when compared with the overall population):

- Are younger
- Are either currently studying in a college or are employed with graduate education.
- Comes from developed countries, especially from countries where math skills is not that highly rated

For the pass rate, we noticed that all sub-populations perform worse in a self-paced course than instructor-paced format. The overall rate pass in the self-paced course has been 15% with the range from -11% to -20%. The relationship between demographic features and the pass rate is more intricate with not very clear interpretation. Countries play an important role with learners from Spain performing relatively well with the self-paced course as compared to learners from Egypt. Nevertheless, our analysis shows that the groups with the following demographic features perform the worst when it comes to pass rate:

- Includes a higher proportion of students
- Have lower skill, defined by the percentage of correct problems per submission
- Are from countries with moderate English index (between 60 and 70).

8.4 Conclusion

In this chapter, we explored the impact of pacing on students' performance. The transition to self-pacing caused all student success metrics to drop even when one controls for demographics, timing, and skill changes. The completion rate dropped by 6%, pass rate reduced by 10% and engagement by 7%. However, the difference is not uniform across all groups, with some experiencing no change and others encountering an even larger decrease. Older students with more familiarity with the topic and those that have higher math skills performed the same across both the format while college-going students and those from countries with lower math skills performed the worst.

Chapter 9

Perception Measurement: Impact of course pacing on students' perception

In this section, we examine student perception of the course design change. Specifically, we are interested in comparing student attitude towards learning between instructor and self-paced run of the course.

Perception Measures examines students increase in interest in the subject, interpersonal outcomes (e.g., cooperative abilities), intrapersonal outcomes (e.g., self- understanding), and other broad course outcomes [38, 12]. Even though we have taken a broad definition of the success metrics, none of the quantitative metrics can truly tell us what motivates learners and what happens in their head during the course. Student perceptions of learning have also been found to be much highly correlated with the overall ratings of teaching effectiveness [64, 11]. We use the exit survey to capture the student perception of the change in course.

9.1 Interest in the subject

One of the best ways to assess student attitude towards learning is to determine her increase in the interest of the subject. To measure the interest, we used both direct and indirect methods. We asked students a couple of question in the exit survey to gauge their interest in the course. We asked them to quantify how likely they are to enroll in the next class of this program. On similar lines, we asked if the course met their expectation and if they will be willing to recommend this course to their friend. The mean difference and a one-way

Table 9.1: Students' perception of interest in the subject. A one-way analysis of variance (ANOVA) is performed on each dimension. The table reports the statistical and practical significances of the comparison. The former is reported in terms of p-value and the later in terms of Cohen's d value.

	Instructor	**Student**	**p-value**	**Effect Size(d)**
survey.Course_Learning[1]	3.353	3.265	0.013	-0.125
survey.Enroll_Future_Course[2]	4.645	4.580	0.074	-0.090
survey.Recommend_Friend[2]	4.499	4.433	0.098	-0.083
Actually end up taking SC1x	69.9%	69.0%	0.589	0.019

[1] Four-point Likert Scale. [2] Five-point Likert Scale

analysis of variance (ANOVA) statistical test between instructor-paced and self-paced run of the course are shown in the table 9.1. The responses to these questions produced the same results in both the self-paced and instructor-paced course. While the mean is higher in the instructor-paced course(and is statistical significant too), the effect is too small and can be ignored. Students engaged very differently in the self-paced course and even performed poorly than the instructor-paced format, but surprisingly, both the groups expressed an equal amount of excitement in the subject.

An indirect way to judge student interest is to observe if they actually enroll in the next course in the program, SC1x. We tracked students enrollment in SC1x and reported the results in the table. The table shows the average number of passing students that enrolled in SC1x for both the different course delivery modes. Surprisingly, the ratio is the same, even though we computed the mean on five different runs of instructor-paced course. Considering that SC0x is the entrance point to the Micromaster program, it is heartening to see that a shift to self-pacing did not reduce the size of the incoming population.

9.2 Interpersonal outcomes

Interpersonal outcomes relate to relationships or communication between people. To measure students' intrapersonal outcome, we asked them to rate their connectedness with the community of learners and instructors in the course. The mean difference and a one-way analysis of variance (ANOVA) statistical test between instructor-paced and self-paced run of the course are shown in the table 9.2. To our surprise, students have rated the connect-

Table 9.2: Students' perception of interpersonal gains from the course. A one-way analysis of variance (ANOVA) is performed on each dimension. The table reports the statistical and practical significances of the comparison. The former is reported in terms of p-value and the later in terms of Cohen's d value.

	Instructor	Student	p-value	Effect Size(d)
survey.Connectedness[1]	2.427	2.561	0.006	0.137
survey.Importance_forums[1]	2.962	3.182	0.000	0.182
survey.Accessiblility[1]	3.19	3.220	0.574	0.028

[1] Five-point Likert Scale

edness higher in self-paced then the instructor-paced version of the course. The difference is statistically significant, although the effect size is minimal. Students also rated the quality of discussion and in general, the importance of forum to the course higher than the instructor-paced format. Although students moved at a different pace and accessed forums at different times, these differences did not let to any degradation in the quality of the forum discussion. These are encouraging results for course designers considering a self-paced format for MOOCs. To conclude, we did not observe any negative impact on Interpersonal communication as a result of the shift to a self-paced version of the course.

9.3 Intrapersonal outcomes

Intrapersonal outcomes happen when a learner internalize a concept. To measure student intrapersonal outcome, we asked them to assess how much they have learned in the course. The mean difference and a one-way analysis of variance (ANOVA) statistical test between instructor-paced and self-paced run of the course are shown in the table 9.3. Students in self-paced learning perceived that they learned slightly less than their counterparts in an instructor-paced version. However the result is not statistically significant and can is ignored. Overall, student across both the formats perceived the gain in intrapersonal outcomes due to SC0x mostly the same.

9.4 Course Content

Next, we examine the student perception of the quality of the course and its components. To address this, we asked the student to rate the quality, on a 5 points scale, of the course

Table 9.3: Students' perception of intrapersonal gains from the course. A one-way analysis of variance (ANOVA) is performed on each dimension. The table reports the statistical and practical significances of the comparison. The former is reported in terms of p-value and the later in terms of Cohen's d value.

	Instructor	**Student**	**p-value**	**Effect Size(d)**
survey.Course_Learning[1]	3.353	3.265	0.013	-0.125
survey.Course_Expectation[1]	2.768	2.723	0.253	-0.058

[1] Four-point Likert Scale

Table 9.4: Students' perception of quality of the course. A one-way analysis of variance (ANOVA) is performed on each dimension. The table reports the statistical and practical significances of the comparison. The former is reported in terms of p-value and the later in terms of Cohen's d value.

	Instructor	**Student**	**p-value**	**Effect Size(d)**
survey.Quality_Videos[1]	4.271	4.247	0.544	-0.031
survey.Quality_Platform[1]	4.207	4.057	0.000	-0.190
survey.Quality_Instructions[1]	4.138	4.071	0.112	-0.081
survey.Quality_QQ[1]	3.925	4.058	0.002	0.155
survey.Quality_PP[1]	3.962	3.870	0.033	-0.108
survey.Quality_Forum[1]	3.510	3.646	0.007	0.136

[1] Five-point Likert Scale

and its core components: instruction, platform, videos, practice problems, quick questions, and forums. The mean difference and a one-way analysis of variance (ANOVA) statistical test between instructor-paced and self-paced run of the course are shown in the table 9.4. Students rated quick questions and forums higher in self-paced. The difference is statistically significant, but the effect size is small. All other components were rated lower, but the difference is statistically significant only for platform quality with a moderate effect size. While the effect size is minimal, this result points towards possible action that can be taken to improve the quality of the EdX components as well as the assessment in the future self-paced run of the course. Perhaps the initial focus of EdX as well as of MITx on instructor-based course delivery created a platform and content that does not seamlessly translate to course delivery based on self-pacing. However, further research is required on a different course to make any strong claim on this subject.

Table 9.5: Students' reasons for pursuing verified certificate. A chi-square test is performed on each dimension. The table reports the statistical and practical significances of the comparison. The former is reported in terms of p-value and the later in terms of phi coefficient.

	Instructor	**Student**	**p-value**	**Effect Size(ϕ)**
survey.academic_advancement[1]	0.439	0.389	0.020	0.054
survey.career_advancement[1]	0.709	0.652	0.004	0.067
survey.networking[1]	0.030	0.050	0.005	0.065
survey.personal development[1]	0.697	0.619	0.000	0.092
survey.professional_certification[1]	0.519	0.473	0.700	0.009

[1] Yes/No response

9.5 Certification

Finally, we examine the student self-reported reason for pursuing the certificate. The mean difference and a chi-square statistical test between instructor-paced and self-paced run of the course are shown in the table 9.5. We observe a relatively moderate difference between the two groups. However, the effect size is small, implying that there is little association between course mode delivery and the certification intent.

9.6 Conclusion

In this chapter, we explored the impact of pacing on students' attitude towards the course. Student perception of learning and their satisfaction with the course remained mostly the same across both the format with some minimal changes. Students expressed the same amount of interest in the subject after the course with an equal proportion of passing student enrolling in the next course of the program. Surprisingly, more students in a self-paced course rated intrapersonal outcome higher than the instructor-paced course, indicating that the self-pacing did not impact the quality of forum discussion. However, students in self-paced format did rank their satisfaction with platform and problems content slightly lower, suggesting that there are opportunities to redesign the content that was originally created for the instructor-paced courses.

Chapter 10

Impact of course pacing on predictive models

In this chapter, we briefly investigate the impact of the course pacing on the predictive model. We focus only on completion metric, although the method described here is generalizable to other success metrics. Specifically, we are interested in answering the following two questions: 1) Can we predict dropouts in a self-paced course with a model pre-trained with data from instructor-paced course?. 2) Do the factors impacting dropout remains the same in both the versions?

Both of these questions are related. However, our focus in 1 is to determine the accuracy of the predictive model and in 2 is to comment on the features that contribute the most in these models.

10.1 Problem Formulation

Given we are at week w (end of week), we need to determine if a user u will drop out or not. Dropouts are those that do not attempt the final exams in the 12th week from the start of course.

We formulate this as a sequence pair $\{(X_i^t, X_i^c, y_i)\}_{i=1..N}$ where each X_i^t is a multivariate sequence (e.g. number of video watched, problems attempted, etc) with T_i samples $< x_{i,1}, x_{i,2}, ..., x_{i,T} >$; X^c is contextual data associated with each sequence(e.g. gender, education, etc); and y is a categorical class label (dropout or complete). Give this information, we seek a function such as

$$h_w : (X^t, X^c)_w \to y_w \tag{10.1}$$

where w goes from 1 to 11 and X_i^t is a multivariate sequence with w samples. In other words, we build a model for each week with input as data from course start to that week and output as a binary categorical variable.

10.2 Method

We then proceeded as follows [16, 17]:

- A sliding window was applied to the sequence data to transform it into a piecewise representation (segments). (width=2, overlap=0.5)
- For each segment, we computed the following features: mean, variance, standard deviation, maximum, minimum, skewness, kurtosis, mean crossings, mean spectral energy, and a 4-bin histogram.
- Features were scaled to zero mean and norm standard deviation, and a Gini importance ranking, feature selection strategy, was implemented to reduce the feature space.
- At this stage, contextual data was appended with the sequential data. We also identified highly correlated features and removed them.
- The data was split into train and test where we used train data to build a model and test data to validate the performance of the model (train-test split = 0.8:0.2).
- We then build a model for each week with input as data from course start to that week and output as a binary categorical variable. Several classification algorithms (naive bayes, logistic, CART, Random Forest, Boosted Trees) were explored. Further, we performed 3-fold cross-validation with grid search to fine-tune the hyperparameters for the algorithms listed earlier.
- Finally, we evaluated our models on the test set using the area under the curve(AUC) score.

	Train Data Source	Test Data Source	Wk1 Score	Wk4 Score	Wk8 Score
Baseline Model	Instructor-paced	Instructor-paced	73%	79%	89%
Model 2	Instructor-paced	self-paced	62%	66%	74%
Model 3	self-paced	self-paced	68%	74%	82%

Table 10.1: Taxonomy of predictive model used in our stud

Weeks	Naïve bayes	Decision Trees	Logistic Regression	Random Forest	Boosted Trees
1	63%	65%	69%	70%	73%
2	67%	65%	72%	72%	74%
3	69%	68%	76%	73%	76%
4	58%	72%	78%	77%	79%
5	61%	76%	80%	81%	83%
6	68%	75%	83%	82%	84%
7	68%	82%	87%	86%	87%
8	72% .	86%	89%	89%	89%
9	78%	88%	92%	92%	92%
10	78%	89%	92%	93%	93%
11	82%	93%	95%	96%	96%

Table 10.2: Comparison of Test AUC scores of various classifiers. Each classifier is trained on a data from instructor-paced courses and evaluated on a left-out data from the same courses.

The figure 10-1 and table 10.2 shows the test accuracy (AUC) over weeks for the classifiers explored by us. Boosted Trees performed the best in each of the weeks. We choose Boosted trees to examine research questions posed by us earlier.

10.3 Transferability of predictive model

To answer Q1, we first developed a baseline model using data from the instructor-paced course and then used it to predict dropouts in a self-paced course.

Our baseline model uses test and train data from the instructor- paced version of SC0x. The test and train score is listed in figure 10-2. The model can identify dropouts with an AUC score of 72% in the first week, and the performance increases linearly until it reaches 95% by the end of the course.

With our baseline model ready, we now turned towards answering Q1 i.e., if we train a model on data from instructor-paced versions of SC0x, can we achieve reasonable accuracy on self-paced course. To address this question, we trained our model(model 2) using all of the data from instructor-paced runs of SC0x and then evaluated its performance on a

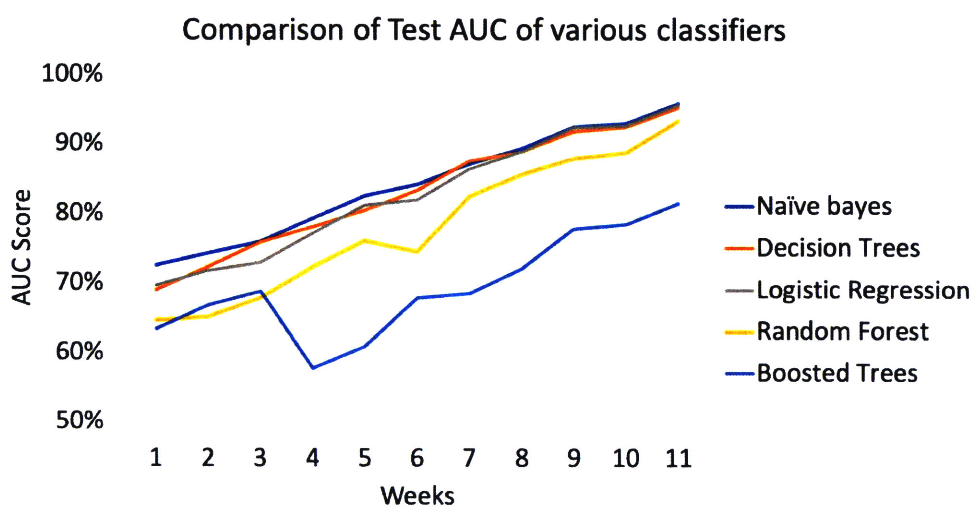


Figure 10-1: Comparison of Test AUC scores of various classifiers. Each classifier is trained on a data from instructor-paced courses and evaluated on a left-out data from the same courses.

self-paced run of SC0x. The test and train score is listed in the figure 10-2. The test score in week 1 drops down to 62% which then improves gradually to 89% by 11th week. The self-paced version of the course has fundamentally changed the students' behavior in the course and a model trained on an instructor-run of the course can no longer predict the dropouts with the same accuracy. This raises questions on the generalization of the predictive model on a course that has undergone design changes.

10.4 Factors correlated with dropouts

To address Q2, we trained another model using data from a self-paced course and then compare the predictors with the baseline model.

To interpret the models, we use SHAP (SHapley Additive exPlanation) values that have been shown as a viable approach to interpret complex and non-linear models. SHAP values explain the output of a function as a sum of the effects of each feature being introduced into a conditional expectation [46]. Since the order in which features are matters, SHAP averages over all possible ordering, giving us a value consistently and accurately attribute the importance of feature on the output.

The figure 10-3 visualizes the SHAP values at week 1, 4 and 8. Every student has one

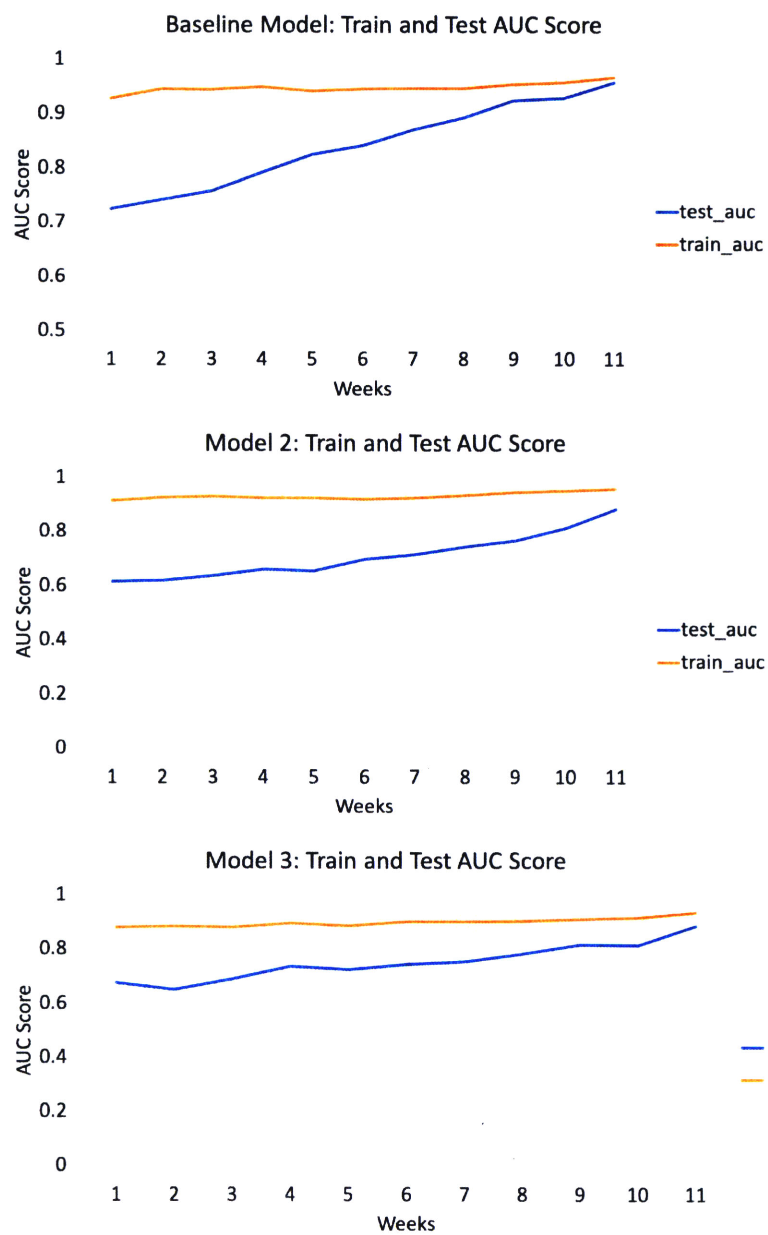


Figure 10-2: Test and Train Score of baseline model (top), model 2 (middle) and model 3 (bottom) over weeks.

dot on each row. The x position of the dot is the impact of that feature on the model's prediction for the student, and the color of the dot represents the value of that feature for the student. The x-axis has units of log-odds [45].

Week 1:

The most important feature in both the models (baseline and model 3) for week 1 is the number of times a student gets the first attempt to the problem correct. Students who get more problems correct on the first attempt have more odds to complete the course. The number of problems submitted and duration of videos watched are also positively correlated with completion and are significant in both the models. We notice that demographic features play a significant role in predicting dropouts in week 1. Verification time, age, and country of residence are essential features in both models. Students who verify earlier have more odds of completing the course. Students in the age bracket between 20 and 30 have more probability of completing the course. US students, on the other hand, are more likely to drop the course. While there are important features that are common in both the models, the magnitude of their impact on the dropouts differs. We also observe features that are important in one model but not in others. For example, in an instructor-paced course, checking course progress page is correlated with a higher completion rate, but this feature does not hold the same importance in a self-paced course. In summary, the features with most predictive power in week 1 are mostly similar across both the models, although the magnitude of their impact on the output differs.

Week 4:

In Week 4, the number of problems where students get the first attempt correct is still one of the most important features across both the models. We do, however, see some major differences in the list of other important features. Activity features have gained importance in the instructor-paced model, while demographic features still have high importance in a self-paced model. For the instructor-paced course, the number of active days and number of progress page visits are positively correlated with completion. Both of these features do not have the same predictive power in a self-paced course. This provides us an explanation of why a model trained with one pacing structure does not lead to good accuracy on a course with a different pacing structure.

Week 8:

Activity features have gained prominence in both the models by week 8. However, the type and magnitude of the impact of these features differ in both the models. The video-based features have more predictive power in the instructor-paced model, while problem-based features enjoy more power in the self-paced model. There are other features that play an important part in one model but not in others. Increased number of visits to the course progress page is correlated with higher completion in the instructor-paced course but does not have the same importance in the self-paced course. On the contrary, problems that are correct on the first attempt still have much importance in the self-paced model but is no longer considered a vital feature in the instructor-paced model.

10.5 Conclusion

In this chapter, we explored the impact of pacing on the predictive model for dropout. Firstly, we investigated if we can use a predictive model trained with one pacing structure on a course with a different pacing format. Our results demonstrate that course pacing fundamentally changes the students' behavior and a model trained on data from the instructor-paced versions of SC0x can no longer predict dropouts in a self-paced version with the same accuracy. Secondly, we examined the factors that are correlated with dropouts in both the different versions of the course. We noticed that both models include similar features(correct on first attempt,active days, videos watched), but the order and magnitude of the impact of the features vary.

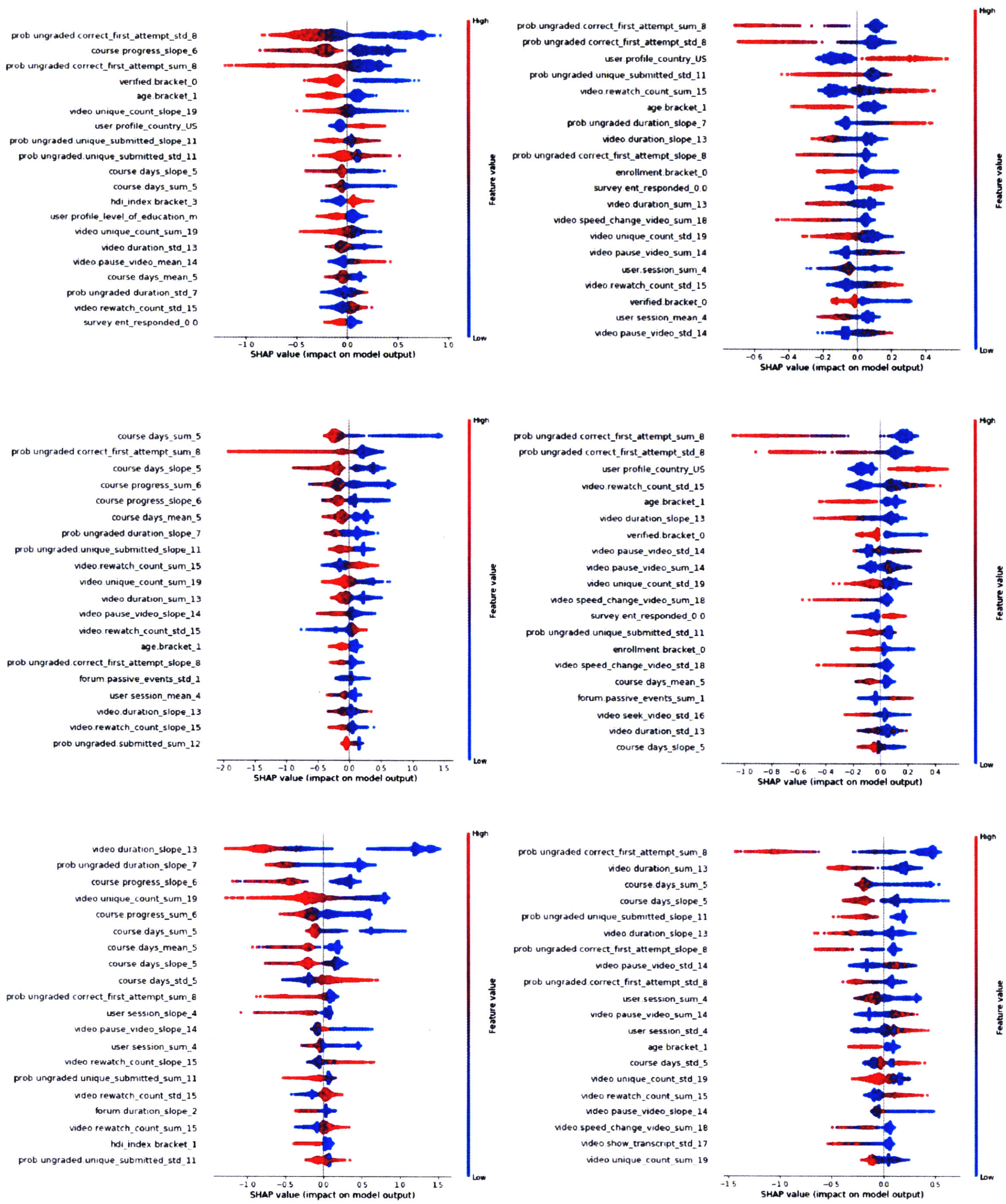


Figure 10-3: Visualization of SHAP values in Week 1 (top), Week 4 (middle) and Week 8 (Bottom) for models trained on data from instructor-paced course (left) and self-paced course (right). In each of the figure every student has one dot on each row. The x position of the dot is the impact of that feature on the model's prediction for the student, and the color of the dot represents the value of that feature for the student. The x-axis has units of log-odds.

Chapter 11

Conclusion, Recommendation and Future Research

In this section, we summarize the key findings of the thesis and provide guidelines for course developers. We also identify the limitations of our research and share possible future

11.1 Conclusion

We begin by concluding our discussion on the framework and then summarize our observations on the application of the framework to evaluate pacing.

11.1.1 Framework to evaluate design changes in MOOCs

In this thesis, we presented a data-driven framework to evaluate complex design changes in MOOCs. We explored a change from multiple angles: process, proficiency, and perception. For process measurement, we integrated clickstream, survey, course meta, and achievement data to examine students' interaction with various course components. The interactions were analyzed from static and temporal viewpoints to understand both the final state and the trajectory that learners took to reach to the final state. For proficiency measurement, we applied causal inference methods to estimate the impact of course design on students' performance. Finally, for perception measurement, we used the exit survey to get students' feedback and their satisfaction with the design change. Through these analyses, we were able to determine the impact of course design on the outcome of students along multiple dimensions of learning and identity aspect of the course where revision is required.

11.1.2 Case Study: Evaluating pacing in MOOCs

We demonstrated the application of this framework by evaluating course pacing on a repeated run of a supply chain MOOC by MITx. The shift to self-pacing caused students to behave differently than those that have undertaken the course with instructor pacing. The most prominent change was noticed in the students' pathway over the duration of the course. In an instructor-paced version, most students proceeded in a similar manner with high engagement at the start, a dip around midterm exam and an upturn in the second half of the course. In a self-paced version, students pursued their own style of trajectory through the course. Some took a cyclical pathway, while others remained uniformly engaged with the material. There were others that started slow and linearly increased their engagement while others started near the end of the course and ramped up their activity exponentially. These contrasting trajectories impacted students' grade, with "uniform" and "linear increase" styles having a positive correlation with course grade. Learners with certain characteristics were inclined to follow a specific pathway with older learners pursuing a "uniform" style and college-going students taking on an "exponential" type of engagement.

While the trajectories are different, the overall count of interactions is mostly similar across the courses with different pacing structure. Students in both pacing format attempted almost the same number of problems and participated in the forums with the same intensity. The video-watching behavior was different with student-pacing students watching far fewer recitation videos than their counterparts in the other pacing format. A more significant difference was, however, seen in the performance of low-achievers/dropouts between the different formats. While the completers interacted similarly in both formats, dropouts engaged less in the self-paced version than those in the instructor-paced course. They watched fewer videos and attempted far fewer problems, with 28% of dropouts not attempting a single practice problem in the self-paced format. In fact, we demonstrated that a predictive model trained on data from instructor-paced versions of SC0x could no longer predict dropouts in a self-paced version with the same accuracy. This validates the change in dropouts' behavior as a result of a shift to self-pacing and raises questions on the generalization of predictive models on a course that has undergone design changes.

The transition to self-pacing caused all student success metrics to drop even when one controls for demographics, timing, and skill changes. The completion rate dropped by 6%, pass rate reduced by 10% and engagement by 7%. However, the difference is not uniform across all groups, with some experiencing no change and others encountering an even larger decrease. Older students with more familiarity with the topic and those that have higher math skills performed the same across both the format while college-going students and those from countries with lower math skills performed the worst.

Student perception of learning and their satisfaction with the course remained mostly the same across both the format with some minimal changes. Students expressed the same amount of interest in the subject after the course with an equal proportion of passing student enrolling in the next course of the program. Surprisingly, more students in a self-paced course rated intrapersonal outcome higher than the instructor-paced course, indicating that self-pacing did not impact the quality of forum discussion. However, students in self-paced format did rank their satisfaction with videos and problems content slightly lower, suggesting that there are opportunities to redesign the content that was originally created for instructor-paced courses.

11.2 Recommendation

From these findings, we provide three recommendations for course developers and faculty to consider while creating or updating their open online courses.

11.2.1 When to use Instructor-pacing?

First, we recommend course developers to consider instructor-paced pacing when deciding on pacing strategy for their MOOCs. We found a causal relationship between pacing and student success metrics with instructor-pacing increasing completion rate, passing rate, and engagement score. These results are in contrast with an earlier study by Reich on a course offered by HarvardX. However, the course they reviewed was qualitative in nature and had minimal assessment structure. Considering the plethora of quantitative courses available on MOOC, our recommendation to course developers that are involved with courses similar to ours is that they should consider an instructor-based pacing strategy for their course.

11.2.2 When to use student-pacing?

Secondly, student-pacing does seem to provide a viable alternative pacing strategy if the instructor is concerned with the students' perception of learning rather than the quantitative metrics. Students rated both types of courses equally when it comes to perception and satisfaction. Previous authors have expressed concern about the impact of asynchronicity on the flow of discussions in the forum. We did not find any drop in forum engagement in the self-paced version of the course. In fact, students spent more time in forums in the self-paced course than in an instructor paced one. Note that we are not making any claim on the impact of discussion flow due to self-pacing. We are just saying that self-pacing did not hinder students from participating in the forums with students rating quality of the discussion same or better than the instructor-paced version.

11.2.3 How to improve success in student-pacing?

Thirdly, for course developers, who do not have the resources to offer an instructor-paced version or have already made the decision to provide an self-paced version of the course, can benefit by implementing the following countermeasures to mitigate the drawback of student-pacing:

- High achievers performed similarly across both the delivery methods but low achievers engaged less with the self-paced version. Course designers must, therefore, introduce changes into the self-paced course to make this subgroup more engaged with the course. A noticeable difference is observed in problem submission of this subgroup. Perhaps, making the problems optional and ungraded reduced learners' motivation to undertake these assessments. In fact, evidence from cognitive psychology supports this hypothesis as testing has been shown to not only assess learning but also facilitates it [48]. We recommend that the course structure of self-paced be changed to incorporate a few graded assessments. The delivery date of these assessments can be set loosely and the number of assignments can be set low so as not to erode the flexibility offered by a self-paced version. Nevertheless, considering how little low achievers engaged with problems, we believe there is a relative advantage to include few graded assignments in the self-paced version of the course. Another approach could be to keep all the problems ungraded but re-frame them to encourage learners to participate in assessment activities. For example, few of the important problems can be picked from the pool of practice problems and grouped into a new category "Recommended problems" to stimulate learners to at least attempt those.

- Asychrnourous pacing impacted the passing rate the most with student-pacing causing the passing rate to decrease by 15% for students who completed the course. A part of the decrease is attributed to a shift in distributing the 100% of the grade to Final Exams. In the instructor-format, the final grade contributed 45% to the overall course grade, while the midterm and the graded assignments contributed to the rest. In a self-paced format, all graded contents were removed except the final exam, and therefore, the final exam contributed 100% to the overall course grade. If we had computed the pass rate only on the final exam score, the overall decrease in the pass rate would have been 10%. There are approximately 5% of completers in a self-paced course that would have passed the course had we kept the grading distribution the same. This gives another data point to support our earlier statement on introducing graded assignments during the duration of the course. Multiple graded contents give an opportunity for students to recover from a bad grade besides keeping them engaged with the course.

- Students behaved differently with video and problem content in the student-pacing version of the course with many activities happening at the near end of the course. The videos and problems in the course were designed with instructor-pacing in mind. Seeing how students interacted with the content in the self-paced course, there are opportunities to alter the content to make it appropriate for courses with self-paced delivery. Perhaps this is why students rated the video and problem lower in satisfaction in a self-paced course. We recommend designers to reconsider the content to align it with the delivery mode of the course better. One possibility that we foresee is to trim the video duration or reduce the video count. Students in self-paced have demonstrated that one can get good grades by watching fewer videos. Reducing the video duration might have a positive impact on dropouts since many students view the percentage of video completed as a benchmark on how well they are doing the course. An increased percentage of video completed might encourage them to participate in the final exam.

- The move to student-pacing provided an opportunity for learners to engage with the course at their own pace and style. We demonstrated that some of the pathways taken by learners are correlated with lower student success metrics. Course designers must, therefore, deploy interventions to avoid learners to take these pathways. A quick and easy fix is to prepare a set of possible study plans and communicate these with the learners. However, previous research has shown that even in course with the recommended syllabus, learners engage with the course at their own pace [52]. Another approach is to deploy a predictive model that can identify types of learners and then intervene by sending targeted emails. Previous research has found emails to be not effective in nudging students towards the desired state, although most studies have not implemented targeted messaging [7]. Another approach can be to make design changes to reduce the impact of various pathways on the success metrics. For example, one could breakdown the course into four sub-module with each module have its own final exam. Reducing the duration of the course might minimize the impact of the pathway on the success metric and also offers an opportunity for the learners to correct their study plan between the sub-modules.

- MOOCs are criticized for low completion rate with six years of research in this space, offering limited practical guidance on how to tackle this problem. Our analysis provides some guidance to researchers and course designers on how one can approach such a complicated problem. MOOCs include a diverse group of learners with distinct needs. A one-fits-all approach to address dropout is more likely not to produce the required results. We demonstrated that one could use complex design experiments to understand the subgroups of learners better. Specifically, a change in a course offers an opportunity to model how a certain group of learners responds to change. We showed possible ways to deconstruct these learners in both static and dynamic context. Understanding these subgroups can lead to better design of interventions. To start with, researchers can develop predictive models that can identify types of dropouts. And, then can deploy interventions specific to that group.

11.3 Limitations and Future Research

There are limitations as well as opportunities to further expand on the work done by us. In this section, we pen down these limitations and provide guidance for future research.

Firstly, while we have applied the framework to evaluate course pacing in MOOC, no single study can validate the effectiveness of the framework. There is a full spectrum of design decisions in MOOCs with varying effort and impact. These changes might be related to a component of a MOOC (such as forum re-design) or might have far-reaching consequences that go beyond online education (such as blended programs). Considering how different and broad these changes might be, there is a legitimate question if the proposed framework will be able to generalize well. We recommend carrying out future research in this direction to validate and improve the efficacy and reach of the framework proposed in this thesis.

Secondly, while our framework provides a systematic approach to identify areas for improvement, it offers limited guidelines to what could be done to improve these areas. Even in situations, where we have provided recommendations, there are opportunities to develop an analytical model to estimate the impact of these proposed changes on student success. Such an approach would prevent instructors from implementing costly interventions that can only have minimal effects on course.

In this thesis, we have proposed an engagement score that captures student's learning in the course. The effectiveness of this score needs to be investigated in other courses of different types and structure. Importantly, however, it needs to be validated if this score indeed measures learning in true sense. While we never might be able to quantify learning, there is a real need to define and operationalize metrics that goes beyond grades or mere clicking. Future research in this direction would have immense value to the larger MOOC community.

The course we investigated had undergone only one design change. However, one is very likely to see a situation where a course had experienced multiple design changes, either separated by time or applied all at once. Consider a scenario, where new assignment types were introduced in a certain week, and the structure of the final exam was changed. The methods explored in this thesis are not comprehensive to evaluate a scenario like this. Methods such as causal methods with time-varying treatments and reinforcement learning have shown positive results on similar scenarios in applications other than MOOC. We propose that future research must be conducted to examine this, and other methods be explored to evaluate multiple design change in MOOCs.

There are opportunities to standardize the data processing pipeline of MOOCs to enable researchers to compare MOOCs across different platforms. To be able to identify design changes that work, one needs to explore data from hundreds of MOOCs available on multiple platforms. Considering that data aggregation pipeline varies substantially across these platforms, course designers can immensely benefit from a framework that can normalize the data collection process. While there have been previous attempts in this direction, further research is required to develop a practical and usable framework to address this issue.

Finally, we want to comment on the limitations of our study on pacing in MOOCs. Firstly, we restate our earlier statement that the observations from this study can not be generalized. Although we have data from six-course, only one of them was self-paced. Additionally, we explored self-pacing only on one type of course. The observations and recommendations from our analysis are directed toward the MITx teaching team. And while it does provide guidance to faculty and course designers (especially to the ones that are involved with the quantitative course like ours), there is no guarantee that the results will generalize to other courses. In fact, previous work on pacing has produced different results. To be able to draw general inference about the impact of course pacing on student success, there is a need to conduct studies using data from multiple MOOC course. Academic universities and MOOC providers should collaborate to provide data to support such efforts.

Secondly, we only considered verified learners and did not measure the impact of pacing on audit learners. While audit learners engage very little with the course, they represent a large group of students. It would have been interesting to see the impact of pacing on their behavior and activities. Considering the size of audit learners in MOOCs, the result would have provided guidance to many course designers looking for ways to engage audit students. Additionally, the teaching team at MITx had given SC2x (an advance course in the Micromaster series) learners audit access to the self-paced SC0x to enable them to revise mathematical concepts. It would have been interesting to see how many of these learners interacted with the course and what has been the impact on audit access on performance in SC2x. Further research must be conducted to explores these questions.

Appendix A

Tables

stratum	Reference	0	1	2	3	4	5	6
# of learners	4409	295	294	588	1175	1175	588	294
% Learners	100%	7%	7%	13%	27%	27%	13%	7%
ATE	-7%	-10%	-12%	-5%	-9%	-7%	-2%	11%
% ATE difference from total population	0%	-49%	-73%	22%	-39%	-3%	70%	264%
gender· female	0 24	(0.17,0.09)***	(0 21,0 04)**	(0 27,0 04)**	(0 35,0 12)***	(0 33,0 1)**	(0 34,0 11)***	(0 27,0 03)*
gender male	0 76	(0.83,0.09)***	(0 79,0.04)**	(0.72,0 04)**	(0 64,0 12)***	(0 67,0 09)**	(0 66,0 11)***	(0 73,0 04)*
edu associate	0 03	(0 03,0 0)	(0 03,0.0)	(0 03,0 0)	(0 04,0 01)	(0 05,0.03)*	(0 05,0 03)**	(0 02,0 01)
edu bachelor	0.51	(0.45,0 09)***	(0 45,0.09)***	(0 55,0.05)**	(0 63,0.11)**	(0 67,0.16)***	(0.68,0 17)***	(0 56,0 05)**
edu· high school	0.06	(0 1,0 05)***	(0 05,0 02)*	(0 03,0.03)***	(0 05,0.01)	(0 05,0 0)	(0 04,0 02)	(0 06,0 0)
edu: masters	0.38	(0 41,0 05)**	(0 46,0 12)***	(0 36,0 02)	(0 26,0.12)**	(0 18,0.19)***	(0 2,0 18)***	(0 3,0 08)***
country: AU	0.02	(0 03,0 01)	(0 03,0.01)**	(0 02,0 01)	(0 02,0 0)	(0 01,0 01)	(0 02,0 01)	(0 01,0 02)**
country· BR	0.05	(0 04,0 01)	(0 02,0 03)***	(0 03,0 03)***	(0 05,0 0)	(0 05,0.0)	(0 12,0 07)***	(0 11,0 07)***
country· CA	0 03	(0.03,0 0)	(0 04,0 02)**	(0 03,0 0)	(0.04,0 01)	(0 02,0 01)	(0.01,0 02)**	(0 02,0 02)*
country· CN	0.03	(0.05,0.02)**	(0 04,0 0)	(0 03,0 0)	(0 02,0 01)	(0 03,0 0)	(0.02,0 01)	(0.03,0 01)
country EG	0 03	(0 04,0 01)**	(0 01,0.02)***	(0 01,0 02)***	(0 01,0 01)	(0.02,0 01)	(0.02,0 01)	(0 09,0 07)***
country ES	0.03	(0.05,0 03)***	(0 02,0.01)	(0 02,0 02)**	(0.03,0 0)	(0.05,0 02)	(0.03,0 0)	(0 01,0 02)**
country IN	0.12	(0 08,0 05)***	(0.11,0.02)	(0 12,0 0)	(0 21,0 09)***	(0.14,0 02)	(0.21,0 09)***	(0 15,0.04)**
country: MX	0.03	(0 03,0 0)	(0.03,0.0)	(0 03,0 01)	(0 01,0 02)	(0.05,0 02)	(0.01,0 02)	(0.02,0 01)
country_US	0 32	(0 32,0 0)	(0 35,0 04)*	(0 33,0 02)	(0.33,0 01)	(0 37,0 05)	(0.3,0.02)	(0.24,0.09)***
os· Mac OS X	0 12	(0 21,0 12)***	(0 16,0 05)***	(0 06,0 08)***	(0 03,0 1)***	(0 05,0 08)**	(0 05,0.07)***	(0 07,0 06)***
os Windows	0 87	(0.77,0.13)***	(0 82,0 06)***	(0 93,0.09)***	(0 97,0.11)***	(0 95,0 08)**	(0 94,0.08)***	(0 92,0.06)***
age	33 07	(31 41,0.22)***	(32 87,0.05)**	(33.85,0 12)***	(35 0,0.18)***	(35 11,0.18)***	(34 25,0 12)***	(33.6,0 1)***
enrollment date	32.11	(64 14,0 52)***	(27.45,0.02)	(16 53,0 24)***	(12 82,0 31)***	(8 44,0 36)***	(11 91,0 35)***	(29 22,0 16)***
verified time	-1 94	(-0 07,0 14)***	(-0 07,0 08)***	(-0 03,0 08)***	(-2 13,0 02)	(-5 36,0.07)	(-3 76,0 09)**	(-11 48,0 44)***
joining date	443 93	(415.44,0 1)***	(351.02,0.08)***	(375 34,0.13)***	(496 48,0 0)	(518 76,0.02)	(631 88,0 06)	(701.42,0 17)***
english_index	73 12	(72 59,0 02)	(75 09,0 08)***	(73 99,0 03)*	(74 98,0 07)	(74 21,0.02)	(71 23,0 02)	(67 35,0.16)***
hdi	0.83	(0 84,0 05)**	(0 84,0.09)***	(0 83,0 01)	(0 8,0.1)*	(0.83,0 01)	(0.8,0 14)***	(0 79,0 17)***
skill	0 74	(0.71,0 04)*	(0.77,0.14)***	(0.74,0.09)***	(0 74,0 08)*	(0.72,0 13)**	(0 74,0 1)**	(0 75,0 04)*
empstatus Employed	0 80	(0.76,0 06)***	(0 8,0 0)	(0 83,0 05)**	(0 84,0 04)	(0.8,0 0)	(0.76,0 04)	(0 83,0 04)
empstatus Full-time student	0.09	(0.15,0.08)***	(0.1,0 01)	(0 06,0.05)***	(0 05,0 04)	(0.04,0 06)*	(0 08,0 01)	(0.05,0.05)**
empstatus· Other	0.03	(0 03,0.0)	(0 03,0 01)	(0 03,0.0)	(0 03,0 0)	(0 01,0.02)	(0.02,0 01)	(0 04,0 01)
empstatus· Retired	0 00	(0 0,0.0)	(0 0,0 0)	(0 0,0 0)	(0.0,0 0)	(0 0,0 0)	(0.0,0 0)	(0 0,0 0)
empstatus: Unemployed	0.08	(0 06,0 03)*	(0.08,0.0)	(0 08,0 0)	(0 08,0.0)	(0 15,0 07)**	(0 14,0 06)**	(0 08,0 0)
fam: Not at all familiar	0 05	(0 04,0 01)	(0 05,0.0)	(0 06,0 01)	(0 07,0 02)	(0 09,0 04)	(0.06,0 01)	(0.04,0 01)
fam· Slightly familiar	0 16	(0 14,0 02)	(0 15,0.0)	(0 17,0 02)	(0.14,0.01)	(0 12,0 03)	(0 19,0 03)	(0 17,0 01)
fam Somewhat familiar	0 50	(0.53,0 04)*	(0.49,0 01)	(0 47,0 04)*	(0 62,0 12)*	(0 55,0 05)	(0 48,0 02)	(0.48,0 02)
fam· Very familiar	0 25	(0.24,0 01)	(0.25,0 01)	(0 27,0 03)	(0 14,0 1)*	(0 22,0 02)	(0.24,0.01)	(0.23,0 01)
fam Extremely familiar	0.05	(0.05,0.01)	(0 05,0 0)	(0 04,0 01)	(0.03,0 02)	(0 01,0.04)	(0 04,0.01)	(0 08,0.03)**
fluency· Basic	0 01	(0 01,0 0)	(0 01,0 0)	(0 01,0 0)	(0 01,0 01)	(0 0,0 01)	(0 0,0 01)	(0 0,0 01)
fluency Fluent	0 58	(0 58.0 0)	(0 58 0 0)	(0 59,0 02)	(0 57,0 01)	(0 69,0 11)*	(0.53,0 05)	(0 54,0 04)
fluency Intermediate	0 11	(0 11 0 01)	(0 12 0 01)	(0 1 0 01)	(0 12,0 01)	(0 08,0 04)	(0 13,0 02)	(0 12,0 01)
fluency Proficient	0 30	(0 3,0 0)	(0 3,0 01)	(0 3,0 01)	(0 3,0 0)	(0 24,0 07)	(0 34,0 03)	(0 33,0 03)
fluency Week	0.00	(0 0,0 0)	(0 0,0 0)	(0 0,0 0)	(0 0,0 0)	(0 0,0 0)	(0 0,0 0)	(0 0,0 0)**

Table A.1: Clusters and their demographic profile for completion metrics. A Mann-Whitney U test is performed between each cluster and the overall population minus that cluster. The output is displayed as (a, b) where a is the percentage mean difference and b is the effect size.

stratum	Reference	0	1	2	3	4	5	6	7	8
Abs Learners	2754	178	177	354	708	718	355	175	89	88
% Learners	100%	3%	3%	7%	14%	14%	7%	3%	2%	2%
ATE	-10%	-3%	-6%	-16%	-12%	-11%	-9%	-14%	-13%	-6%
% ATE difference from total population	0%	73%	44%	-58%	-19%	-10%	12%	-38%	-24%	37%
gender female	0 239322	(0 27,0 03)	(0 27,0 03)	(0 27,0.04)	(0 25,0 02)	(0 21,0 04)*	(0 21,0.03)	(0 25,0 01)	(0 21,0 03)	(0 23,0 01)
gender male	0 757854	(0 73,0 03)	(0 73,0 03)	(0 72,0 04)*	(0 74,0 02)	(0 79,0.04)*	(0 79,0 04)	(0 75,0 01)	(0 79,0 03)	(0 77,0 02)
edu. associate	0 021885	(0 01,0 02)	(0 02,0 0)	(0 02,0 01)	(0 02,0 0)	(0 03,0 0)	(0 03,0 01)	(0 03,0 01)	(0 01,0 01)	(0 03,0 01)
edu bachelor	0 519238	(0 37,0 16)***	(0 4,0 13)***	(0 42,0 12)***	(0 51,0 02)	(0 51,0 01)	(0 64,0 14)***	(0 68,0 17)***	(0 65,0 14)**	(0 67,0 16)**
edu high school	0 051182	(0 21,0 17)***	(0 1,0 05)***	(0 08,0 03)**	(0 03,0 03)**	(0 03,0 03)***	(0 02,0 03)**	(0 04,0 01)	(0 02,0 03)	(0 06,0 01)
edu masters	0 38581	(0 39,0 0)	(0.45,0 07)*	(0 48,0 11)***	(0 43,0.06)**	(0 41,0 03)	(0 27,0 13)***	(0 23,0 17)***	(0 27,0 12)*	(0 23,0 16)***
country AU	0 027886	(0 04,0 01)	(0 02,0 01)	(0 0,0 03)**	(0 01,0 03)***	(0 01,0 02)***	(0 05,0 02)**	(0 09,0 07)***	(0 18,0 16)***	(0 08,0 05)**
country BR	0 054359	(0 01,0 05)**	(0 01,0 05)**	(0 03,0 03)*	(0 03,0.04)***	(0 06,0 0)	(0 11,0 06)***	(0 16,0 11)***	(0 12,0 07)**	(0 05,0 01)
country CA	0 030709	(0 01,0 03)*	(0 02,0 01)	(0 05,0 02)*	(0 05,0 02)**	(0 03,0 0)	(0 02,0 01)	(0 02,0 01)	(0 0,0 03)*	(0 0,0 03)*
country CN	0 038475	(0 05,0 01)	(0.06,0 02)	(0 05,0 01)	(0 04,0 01)	(0 04,0 0)	(0 04,0 0)	(0 02,0 02)	(0 0,0 04)*	(0 0,0 04)*
country EG	0 034239	(0 1,0 07)***	(0 02,0 01)	(0 03,0 01)	(0 01,0.03)***	(0 01,0 03)***	(0 01,0 03)**	(0 03,0 0)	(0 06,0 02)	(0 43,0 41)***
country ES	0 032121	(0 22,0 2)***	(0 04,0 01)	(0 05,0 02)*	(0 01,0.02)***	(0 02,0 02)**	(0 01,0 03)**	(0 01,0 03)*	(0 01,0 02)	(0 0,0 03)*
country IN	0 127074	(0 09,0 04)	(0 12,0 0)	(0 12,0 01)	(0 11,0.02)	(0 14,0 02)	(0 16,0 04)*	(0 16,0 03)	(0 09,0 04)	(0 08,0 05)
country MX	0 031768	(0 01,0 03)*	(0 03,0 0)	(0 02,0 02)*	(0 03,0 0)	(0 04,0 01)	(0 04,0 01)	(0 04,0 01)	(0 04,0 01)	(0 01,0 02)
country_US	0 268973	(0 23,0 04)	(0 33,0 06)*	(0 4,0 15)***	(0 35,0 11)***	(0 25,0 02)	(0 14,0 15)***	(0 16,0 11)***	(0 12,0.15)***	(0 05,0 23)***
os Mac OS X	0 155312	(0 17,0 01)	(0 44,0 3)***	(0 31,0 17)***	(0 22,0 09)***	(0 04,0 15)***	(0 04,0.13)***	(0 05,0 12)***	(0 03,0 13)***	(0 1,0 05)
os Windows	0 830921	(0 79,0 05)	(0 55,0 3)***	(0 67,0 18)***	(0 76,0.1)***	(0 95,0 16)***	(0 95,0.14)***	(0 95,0 13)***	(0 96,0 13)***	(0 88,0 05)
age	32 395351	(27 97,0 51)***	(27 73,0 49)***	(30 31,0 22)***	(33 02,0 07)**	(34 56,0 26)***	(34 11,0 21)***	(32 08,0 02)	(30 89,0 09)	(31 89,0.01)
enrollment date	31 269326	(66 4,0 43)***	(53 63,0 33)***	(50 95,0 32)***	(37 82,0 18)***	(22.77,0 14)***	(13 27,0 3)***	(6 14,0 44)***	(-2 76,0 55)***	(9 03,0 4)***
verified time	-7 770914	(2 04,0 1)**	(-2 52,0 07)	(2 0,0 13)***	(-4 22,0 09)***	(-9 3,0 02)	(-15 27,0 13)***	(-19 64,0 24)***	(-19 81,0 26)***	(-27 58,0 35)***
joining date	436 745146	(532 71,0 21)***	(396 8,0 04)	(348 96,0 0)	(310 45,0 15)***	(346 53,0 14)***	(549 3,0 08)**	(806 06,0 28)***	(922 55,0 33)***	(734 0,0 13)*
english_index	70 751405	(66 78,0 11)**	(74 5,0 1)*	(76 63,0 17)***	(74 38,0 13)***	(69 9,0 0)	(64 96,0 14)***	(68 26,0 09)*	(67 92,0 14)*	(56 4,0 55)***
hdi	0 817147	(0 82,0 04)	(0 83,0 07)	(0 84,0.12)***	(0 83,0 09)***	(0 82,0 01)	(0 79,0 13)***	(0 79,0 1)*	(0 81,0 0)	(0 74,0 38)***
skill	0 742067	(0 74,0 01)	(0 76,0 02)	(0 72,0 08)*	(0 74,0 01)	(0 74,0 0)	(0 75,0 02)	(0 77,0 07)*	(0 76,0 03)	(0 73,0 04)
empstatus Employed	0 793497	(0 57,0 24)***	(0 74,0 06)	(0 76,0 04)	(0 83,0 05)*	(0 84,0 07)**	(0 86,0 08)**	(0 72,0 08)*	(0 69,0 11)*	(0 7,0 09)
empstatus Full-time student	0 094695	(0 3,0 22)***	(0 17,0 08)**	(0 12,0 03)	(0 06,0 05)***	(0 06,0 05)**	(0 05,0 06)**	(0 11,0 01)	(0 12,0 03)	(0 16,0 07)
empstatus Other	0 029093	(0 05,0 02)	(0 04,0 02)	(0 03,0 0)	(0 03,0 0)	(0 02,0 01)	(0 03,0 0)	(0 04,0 02)	(0 02,0 01)	(0 05,0 02)
empstatus Retired	0 00057	(0 0,0 0)	(0 0,0 0)	(0 0,0 0)**	(0 0,0 0)	(0 0,0 0)	(0 0,0 0)	(0 0,0 0)	(0 0,0 0)	(0 0,0.0)
empstatus Unemployed	0 082145	(0 08,0 0)	(0 05,0 03)	(0 08,0 0)	(0 09,0.01)	(0 08,0 01)	(0 06,0.03)	(0 13,0 05)*	(0 17,0 09)*	(0 09,0 01)
fam Not at all familiar	0 041072	(0 03,0 01)	(0 02,0 03)	(0 06,0 02)	(0 04,0 0)	(0 04,0 0)	(0 05,0.01)	(0 07,0 03)	(0 04,0 0)	(0 0,0 04)
fam Slightly familiar	0 146035	(0 17,0 02)	(0 09,0 06)*	(0 15,0 01)	(0 16,0 02)	(0 13,0 02)	(0 16,0.02)	(0 14,0 0)	(0 17,0 02)	(0 16,0 01)
fam Somewhat familiar	0 495722	(0 5,0 01)	(0 59,0 1)*	(0 51,0 02)	(0 5,0 0)	(0 49,0 01)	(0 45,0.05)	(0 5,0 0)	(0 46,0 04)	(0 45,0 04)
fam Very familiar	0 2664	(0 27,0 0)	(0 27,0 01)	(0 23,0 05)	(0 25,0 02)	(0 29,0.03)	(0 29,0.03)	(0 23,0 04)	(0 31,0 05)	(0 3,0 03)
fam Extremely familiar	0 05077	(0 04,0 02)	(0 04,0 02)	(0 05,0 0)	(0 06,0 01)	(0 05,0.0)	(0 05,0.01)	(0 07,0 02)	(0 02,0 03)	(0 09,0 04)
fluency Basic	0 007416	(0 0,0 01)	(0 01,0 0)	(0 0,0 0)	(0 0,0 0)	(0 01,0 0)	(0 01,0 01)	(0 0,0 01)	(0 0,0 01)	(0 02,0 02)
fluency Fluent	0 54421	(0 44,0 11)*	(0 59,0 05)	(0 65,0 12)***	(0 61,0 08)***	(0 52,0 04)	(0 49,0 07)*	(0 52,0 02)	(0 44,0 11)	(0 27,0 28)***
fluency Intermediate	0 116942	(0 18 0 07)*	(0 08,0 04)	(0 08,0 04)*	(0 08,0 05)**	(0 12,0 01)	(0 17 0 06)**	(0 14 0 03)	(0 15,0 03)	(0 18,0 07)
fluency Proficient	0 331432	(0 38,0 05)	(0 32,0 01)	(0.27,0 07)*	(0 31,0 03)	(0 35,0 03)	(0 33,0 0)	(0 34 0 01)	(0 42,0 09)	(0 52,0 2)**

Table A.2: Clusters and their demographic profile for pass rate metrics. A Mann-Whitney U test is performed between each cluster and the overall population minus that cluster. The output is displayed as (a, b) where a is the percentage mean difference and b is the effect size.

gender/course	20171T	20173T	20181T	20183T	20191T
female	21.9%	23.2%	24.0%	22.8%	25.6%
male	77.8%	76.5%	76.0%	77.1%	74.0%
other	0.2%	0.4%		0.1%	0.4%

Table A.3: learners' gender ratio in all runs of SC0x

age/course	20171T	20173T	20181T	20183T	20191T
0-20	0.1%	0.4%	0.1%	0.4%	0.2%
20-30	38.7%	38.7%	34.8%	34.6%	29.3%
30-40	40.9%	41.3%	48.2%	49.7%	55.2%
40-50	14.6%	14.1%	12.5%	12.1%	10.6%
50+	5.7%	5.4%	4.3%	3.2%	4.7%

Table A.4: learners' age ratio in all runs of SC0x

country/course	20171T	20173T	20181T	20183T	20191T
US	30.9%	32.0%	32.8%	31.6%	29.8%
IN	10.5%	9.8%	9.6%	12.5%	12.6%
BR	6.2%	4.3%	4.2%	3.9%	5.2%
MX	3.4%	3.9%	2.2%	3.7%	2.5%
CN	3.2%	2.3%	4.0%	2.7%	3.7%
EG	2.6%	3.2%	2.2%	2.6%	3.2%
ES	2.5%	2.6%	2.6%	3.2%	2.1%
CA	2.4%	3.3%	2.6%	3.1%	3.1%
DE	2.1%	1.8%	2.2%	1.7%	1.6%
CO	1.2%	2.5%	1.6%	0.9%	1.3%
Others	35.0%	34.4%	36.0%	34.0%	35.0%

Table A.5: Country of residence of learners in all runs of SC0x

user.enrollment_course_id	20171T	20173T	20181T	20183T	20191T
junior high schooll	0.4%	NaN	0.2%	0.1%	0.1%
high school	6.6%	6.0%	6.6%	7.1%	4.8%
associate	2.5%	2.4%	2.5%	2.6%	3.0%
bachelor	50.4%	50.8%	49.6%	50.4%	53.2%
masters	37.5%	39.1%	38.4%	38.0%	36.5%
PhD	2.0%	1.0%	1.9%	1.2%	1.7%
others	0.6%	0.6%	0.8%	0.5%	0.7%

Table A.6: Level of education of learners in all runs of SC0x

hdi(x)/course	20171T	20173T	20181T	20183T	20191T
($x < 50$)	1%	0%	1%	2%	0%
($50 \leq x < 70$)	21%	20%	19%	23%	24%
($70 \leq x < 90$)	31%	31%	29%	28%	29%
($x > 90$)	47%	49%	52%	48%	47%

Table A.7: HDI of country of residence of learners in all runs of SC0x

english_index(x)/course	20171T	20173T	20181T	20183T	20191T
($25 \leq x < 50$)	17%	19%	15%	15%	16%
($50 \leq x < 75$)	43%	39%	42%	43%	44%
($75 \leq x < 100$)	39%	42%	43%	42%	40%

Table A.8: Englisg language index of learners in all runs of SC0x

enrollment date/course	20171T	20173T	20181T	20183T	20191T
30 days before course	39%	39%	44%	36%	36%
14 days before course	56%	54%	57%	49%	48%
before start of course	79%	79%	78%	71%	70%
within 14 days of start of course	95%	95%	95%	95%	86%
within 30 days of start of course	100%	100%	100%	100%	91%
after 30 days of start of course	100%	100%	100%	100%	100%

Table A.9: Enrollment registration of learners in all runs of SC0x

verification date/course	20171T	20173T	20181T	20183T	20191T
30 days before course	9%	10%	12%	13%	18%
14 days before course	16%	16%	17%	20%	24%
before start of course	29%	32%	32%	33%	43%
within 14 days of start of course	45%	53%	52%	56%	63%
within 30 days of start of course	98%	100%	100%	100%	71%
after 30 days of start of course	100%	100%	100%	100%	100%

Table A.10: Verification registration of learners in all runs of SC0x

edx join date/course	20171T	20173T	20181T	20183T	20191T
before 1 year	36%	34%	37%	37%	38%
before 1 month	35%	37%	36%	36%	36%
before start of course	20%	21%	16%	17%	15%
after start of course	9%	9%	10%	10%	11%

Table A.11: EdX joining data of learners in all runs of SC0x

empstatus/course	20171T	20173T	20181T	20183T	20191T
Full-time student	9%	9%	9%	9%	8%
Unemployed	9%	9%	7%	7%	8%
Employed	63%	70%	74%	76%	75%
Retired	0%	0%	0%	0%	0%
Other	2%	2%	3%	3%	3%

Table A.12: Employment status of learners in all runs of SC0x

fluency/course	20171T	20173T	20181T	20183T	20191T
Weak	0%	-	-	-	0%
Basic	0%	0%	0%	1%	1%
Intermediate	9%	9%	12%	9%	10%
Fluent	41%	52%	53%	53%	51%
Proficient	24%	28%	25%	30%	29%

Table A.13: Language proficiency of learners in all runs of SC0x

familiarity/course	20171T	20173T	20181T	20183T	20191T
Not at all familiar	6%	6%	6%	6%	4%
Slightly familiar	13%	15%	16%	15%	13%
Somewhat familiar	38%	42%	46%	48%	46%
Very familiar	22%	24%	22%	21%	26%
Extremely familiar	5%	4%	4%	4%	6%

Table A.14: Familiarity with topic of learners in all runs of SC0x

Bibliography

[1] Principles of course design. `https://resources.depaul.edu/teaching-commons/teaching-guides/course-design/Pages/principles-of-course-design.aspx`. Accessed: 2019-08-01.

[2] G. Abadie, A. Imbens. Large sample properties of matching estimators for average treatment effects. *Econometrica*, 2006.

[3] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. Engaging with massive online courses. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 687–698, New York, NY, USA, 2014. ACM.

[4] Rachel Baker, Brent Evans, and Thomas Dee. A randomized experiment testing the efficacy of a scheduling nudge in a massive open online course (mooc). *AERA Open*, 2(4):2332858416674007, 2016.

[5] Iain Barr. Causal inference with python. `http://www.degeneratestate.org/posts/2018/Mar/24/causal-inference-with-python-part-1-potential-outcomes/`. Accessed: 2019-08-01.

[6] Gregory S. Black, Clayton L. Daughtrey, and Jeffrey S. Lewis. The importance of course design on classroom performance of marketing students. *Marketing Education Review*, 24(3):213–226, 2014.

[7] Inma Borrella, Sergio Caballero-Caballero, and Eva Ponce-Cueto. Predict and intervene: Addressing the dropout problem in a mooc-based program. pages 1–9, 06 2019.

[8] Christopher Brooks, Craig Thompson, and Stephanie Teasley. A time series interaction analysis method for building predictive models of learners using log data. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, LAK '15, pages 126–135, New York, NY, USA, 2015. ACM.

[9] Andreas Bulling, Ulf Blanke, and Bernt Schiele. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, 46(3):33, 2014.

[10] Julia Cambre, Chinmay Kulkarni, Michael S. Bernstein, and Scott R. Klemmer. Talkabout: Small-group discussions in massive global classes. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 161–162, New York, NY, USA, 2014. ACM.

[11] Ronald G. Cashin, William E.; Downey. Using global student rating items for summative evaluation. *Journal of Educational Psychology*, 1992.

[12] John A Centra and Noreen B Gaubatz. Student perceptions of learning and instructional effectiveness in college courses. *Research Rep*, (9), 2000.

[13] Derrick Coetzee, Armando Fox, Marti A. Hearst, and Björn Hartmann. Should your mooc forum use a reputation system? In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, pages 1176–1187, New York, NY, USA, 2014. ACM.

[14] Koller Daphne, N Andrew, D Chuong, and C Zhenghao. Retention and intention in massive open online courses. *Educause Review*, 48(3):62–63, 2013.

[15] Kimberly K Daugherty. Backward course design: making the end the beginning. *American journal of pharmaceutical education*, 70(6):135, 2006.

[16] Cari Whyne David Burns. Seglearn: A python package for learning sequences and time series. *arXiv*, 2018.

[17] Michael Hardisty Cari Whyne Patrick Henry Stewart McLachlin David Burns, Nathan Leung. Shoulder physiotherapy exercise recognition: Machine learning the inertial signals from a smartwatch. *arXiv*, 2018.

[18] Dan Davis, Guanliang Chen, Claudia Hauff, and Geert-Jan Houben. Gauging mooc learners' adherence to the designed learning path. In *EDM*, 2016.

[19] Marcia D. Dixson. Measuring student engagement in the online course: The online student engagement scale (ose). 2015.

[20] edX. What is self-paced or instructor-paced? `https://support.edx.org/hc/en-us/articles/115011202847-What-is-self-paced-or-instructor-paced-Which-is-my-course-`. Accessed: 2019-08-01.

[21] EF EPI. Ef english proficiency index. *India: EF Education First Ltd*, 2013.

[22] Carolyn Fallahi. Using fink's taxonomy in course design. *APS Observer*, 24(7), 2011.

[23] M. Fei and D. Yeung. Temporal models for predicting student dropout in massive open online courses. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 256–263, Nov 2015.

[24] L Dee Fink. The power of course design to increase student engagement and learning. *Peer Review*, 9(1):13–17, 2007.

[25] William W Fisher. Hls1x: Copyrightx course report. 2014.

[26] Thomas Friedman. Revolution hits the universities. *The New York Times*, 26(1):2013, 2013.

[27] Mitchell M. Handelsman, William L. Briggs, Nora Sullivan, and Annette Towler. A measure of college student course engagement. *The Journal of Educational Research*, 98(3):184–192, 2005.

[28] Imbens G. Ridder G. Hirano, K. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 2003.

[29] Kim A. Hosler and Bridget D. Arend. The importance of course design, feedback, and facilitation: student perceptions of the relationship between teaching presence and cognitive presence. *Educational Media International*, 49(3):217–229, 2012.

[30] D. Imbens, G. Rubin. Causal inference in statistics, social, and biomedical sciences: An introduction. cambridge university press. *Econometrica*, 2015.

[31] Lee J. Cronbach. *Course Improvement Through Evaluation*, volume 64, pages 235–247. 01 2002.

[32] Eamonn Keogh and Jessica Lin. Clustering of time-series subsequences is meaningless: implications for previous and future research. *Knowledge and Information Systems*, 8(2):154–177, Aug 2005.

[33] Rene Kizilcec and Christopher Brooks. Diverse Big Data and Randomized Field Experiments in Massive Open Online Courses. pages 211–222. Society for Learning Analytics Research (SoLAR), Alberta, Canada, 1 edition, 2017.

[34] René F. Kizilcec, Kathryn Papadopoulos, and Lalida Sritanyaratana. Showing face in video instruction: Effects on information retention, visual attention, and affect. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 2095–2102, New York, NY, USA, 2014. ACM.

[35] René F. Kizilcec, Mar Pérez-Sanagustín, and Jorge J. Maldonado. Recommending self-regulated learning strategies does not improve performance in a mooc. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*, L@S '16, pages 101–104, New York, NY, USA, 2016. ACM.

[36] René F. Kizilcec, Chris Piech, and Emily Schneider. Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, LAK '13, pages 170–179, New York, NY, USA, 2013. ACM.

[37] René F. Kizilcec, Emily Schneider, Geoffrey L. Cohen, and Daniel A. McFarland. Encouraging forum participation in online courses with collectivist, individualist and neutral motivational framings. 2014.

[38] Jeff Koon and Harry G. Murray. Using multiple outcomes to validate student ratings of overall teacher effectiveness. *The Journal of Higher Education*, 66(1):61–81, 1995.

[39] George D. Kuh. What we're learning about student engagement from nsse: Benchmarks for effective educational practices. *Change: The Magazine of Higher Learning*, 35(2):24–32, 2003.

[40] Tomasz Kusmierczyk and Manuel Gomez-Rodriguez. On the causal effect of badges. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, pages 659–668, Republic and Canton of Geneva, Switzerland, 2018. International World Wide Web Conferences Steering Committee.

[41] Anne Lamb, Jascha Smilack, Andrew Ho, and Justin Reich. Addressing common analytic challenges to randomized experiments in moocs: Attrition and zero-inflation. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, L@S '15, pages 21–30, New York, NY, USA, 2015. ACM.

[42] Doug Lederman, Assessing the Spellings Commission, et al. Inside higher ed. 2014.

[43] T. Warren Liao. Clustering of time series data—a survey. *Pattern Recognition*, 38(11):1857 – 1874, 2005.

[44] Gilles Louppe, Louis Wehenkel, Antonio Sutera, and Pierre Geurts. Understanding variable importances in forests of randomized trees. In *Advances in neural information processing systems*, pages 431–439, 2013.

[45] Scott Lundberg. Interpretable machine learning with xgboost. `https://towardsdatascience.com/interpretable-machine-learning-with-xgboost-9ec80d148d27`. Accessed: 2019-08-01.

[46] Scott M Lundberg, Gabriel G Erion, and Su-In Lee. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*, 2018.

[47] Ignacio Martinez. The effects of nudges on students' effort and performance: Lessons from a mooc. Technical report, Working Paper, EdPolicyWorks, 2014.

[48] Mark A McDaniel, Henry L Roediger, and Kathleen B McDermott. Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic bulletin & review*, 14(2):200–206, 2007.

[49] MIT. Supply chain analytics. `https://www.edx.org/course/supply-chain-analytics-0`. Accessed: 2019-08-01.

[50] MIT. Micromasters program in supply chain management. `https://micromasters.mit.edu/scm/`, Sep 2017. Accessed: 2019-08-01.

[51] Thomas Moore. Why course design is important. `https://tlc.pomona.edu/2010/07/14/discussion-why-course-design-is-important/`. Accessed: 2019-08-01.

[52] Thomas Mullaney and Justin Reich. Staggered versus all-at-once content release in massive open online courses: Evaluating a natural experiment. *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, pages 185–194, 2015.

[53] Robert Murphy, Larry Gallagher, Andrew E Krumm, Jessica Mislevy, and Amy Hafter. Research on the use of khan academy in schools: Research brief. 2014.

[54] Sergiy O. Nesterko, Daniel Seaton, Justin Reich, Joseph McIntyre, Qiuyi Han, Isaac Chuang, and Andrew Ho. Due dates in moocs: Does stricter mean better? In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 193–194, New York, NY, USA, 2014. ACM.

[55] Tertiary Education Quality and Standards Agency. Guidance note: Course design. `https://www.teqsa.gov.au/latest-news/publications/guidance-note-course-design-including-learning-outcomes-and-assessment`, Sep 2017. Accessed: 2019-08-01.

[56] Justin Reich. Mooc completion and retention in the context of student intent. *EDUCAUSE Review Online*, 2014.

[57] Justin Reich. Rebooting mooc research. *Science*, 347(6217):34–35, 2015.

[58] Justin Reich and José A. Ruipérez-Valiente. The mooc pivot. *Science*, 363(6423):130–131, 2019.

[59] Jan Renz, Daniel Hoffmann, Thomas Staubitz, and Christoph Meinel. Using a/b testing in mooc environments. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, LAK '16, pages 304–313, New York, NY, USA, 2016. ACM.

[60] Jan Renz, Thomas Staubitz, Jaqueline Pollack, and Christoph Meinel. Improving the onboarding user experience in moocs. *Proceedings EduLearn*, 2014.

[61] Chin Choo Robinson and Hallett Hullinger. New benchmarks in higher education: Student engagement in online learning. *Journal of Education for Business*, 84(2):101–109, 2008.

[62] David H Rose and Anne Meyer. *Teaching every student in the digital age: Universal design for learning.* ERIC, 2002.

[63] Donald B Rubin. Causal inference using potential outcomes. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

[64] Janice Marie Ryan and Paul David Harrison. The relationship between individual instructional characteristics and the overall assessment of teaching effectiveness across different instructional contexts. *Research in Higher Education*, 36:577–594, 1995.

[65] Alexander O. Savi, Han L. J. van der Maas, and Gunter K. J. Maris. Navigating massive open online courses. *Science*, 347(6225):958–958, 2015.

[66] Tanmay Sinha, Patrick Jermann, Nan Li, and Pierre Dillenbourg. Your click decides your fate: Leveraging clickstream patterns in MOOC videos to infer students' information processing and attrition behavior. *CoRR*, abs/1407.7131, 2014.

[67] Joshua; France-Harris Antoinette Southard, Sheryne; Meddaugh. Can spoc (self-paced online course) live long and prosper? a comparison study of a new species of online course delivery. 2015.

[68] Peter Stokes. The particle accelerator of learning. *Inside Higher Ed*, 22, 2013.

[69] Colin Taylor, Kalyan Veeramachaneni, and Una-May O'Reilly. Likely to stop? predicting stopout in massive open online courses. *CoRR*, abs/1408.3382, 2014.

[70] Sara Topolovec. A comparison of self-paced and instructor-paced online courses: The interactive effects of course delivery mode and student characteristics. 2018.

[71] Kalyan Veeramachaneni, Una-May O'Reilly, and Colin Taylor. Towards feature engineering at scale for data from massive open online courses. *CoRR*, abs/1407.5238, 2014.

[72] W Weeks. Incorporation of active learning strategies in the engineering classroom. In *ASEE Midwest Section Meeting, University of Missouri-Rolla, September*, 2003.

[73] Laurence Wong. Causal inference in python. `http://causalinferenceinpython.org/`, Sep 2017. Accessed: 2019-08-01.