The Influence of Grades on Learning Behavior of Students in MOOCs

by

Li Wang

B.S., Massachusetts Institute of Technology (2018)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Masters of Engineering in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

© Massachusetts Institute of Technology 2019. All rights reserved.

Author .....
Department of Electrical Engineering and Computer Science
May 23, 2019

Certified by .....
Erik Hemberg
Research Scientist
Thesis Supervisor

Certified by .....
Una-May O'Reilly
Principal Research Scientist
Thesis Supervisor

Accepted by .....
Katrina LaCurts
Master of Engineering Thesis Committee

The Influence of Grades on Learning Behavior of Students in MOOCs

by

Li Wang

Submitted to the Department of Electrical Engineering and Computer Science
on May 23, 2019, in partial fulfillment of the
requirements for the degree of
Masters of Engineering in Computer Science and Engineering

Abstract

In this thesis, we explore and analyze impact of grades on the behavior of students in MOOCs (Massive Open Online Courses). MOOCs often use grades to give students feedback on their understanding of course material and to determine whether a student passes the course. To better understand how student behavior is influenced by grade feedback, we conduct a study on the changes of certified students' behavior before and after they have received their grade. We define continuously participating students as students who continuously do graded assignments up until a specific assignment and how their behavior compares to certified students. Afterwards, we look into the effects of past experience and how these metrics can be used to predict grades with various machine learning models. We observe that both certified and continuously participating students do not change their learning behavior after receiving a specific grade. We also observe that both groups of students with lower grades have more consistent learning behavior than those with higher grades. Lastly, we observe no obvious correlation between past experience, student activity, and grade.

Thesis Supervisor: Erik Hemberg
Title: Research Scientist

Thesis Supervisor: Una-May O'Reilly
Title: Principal Research Scientist

Acknowledgments

I would like to thank Erik Hemberg and Una-May O'Reilly for all of the mentorship they have given me throughout my time as a Master's of Engineering Candidate and for the opportunity to do my thesis with ALFA.

I would like to thank my parents Yanshan Wang and Yifang Bai, as well as my brother Robert Wang for the support they have given me throughout my life and especially during my time at MIT.

I would like to thank my high school teacher Ms. Cao for inspiring me to pursue STEM and guiding me to become a student at MIT. I would like to thank Scott Wu for his continued friendship throughout high school and my time at MIT. I would like to thank Karen Fan for supporting me through my time at MIT and for her continued support and friendship.

Contents

1 Introduction **15**

1.1 The Design of MOOCs 15

1.1.1 Certification in MOOCs 16

1.1.2 Grades, Dropout, and Disengagement 17

1.2 Self-Regulated Learning Systems and Feedback 17

1.3 Research Questions 18

2 Background and Related Works **19**

2.1 Performance Prediction 19

2.1.1 Drop Out Prediction 20

2.2 Self-Regulated Learning Feedback Systems 21

3 Data Set and Research Methods **23**

3.1 Certified and Continuously Participating Students Definition 23

3.2 Data Set 24

3.2.1 Course Content 25

3.2.2 Past Experience Data 25

3.3 Overview of Analysis 27

3.3.1 Certified Students Analysis 27

3.3.2 Grade Trajectory Analysis 28

3.3.3 Continuously Participating vs. Certified Students 28

3.4 Key Terms and Definitions for Analysis 28

3.4.1 Grade States 29

3.4.2 Cumulative Activity 29
3.4.3 Delta Activity and Delta Grade 29

4 Certified Student Analysis **31**
4.0.1 Grade Breakdown 31
4.0.2 Activity Analysis 32
4.1 Analysis on Delta Activity 34
4.1.1 Qualitative Analysis of Delta Activity vs. Grade 35
4.2 Comparisons Across Courses and Course Offerings 37
4.2.1 Delta Activity Across Offerings 38
4.2.2 Delta Activity vs. Delta Grade 39
4.3 Discussion 40
4.3.1 Initial Analysis 41
4.3.2 Delta Activity 41
4.3.3 Delta Grade 42
4.4 Past Experience 42
4.5 Summary of Key Observations 43

5 Grade Trajectory Prediction with Machine Learning **45**
5.1 Trajectory and Variable Selection 45
5.2 Model Selection 46
5.3 Hyper-parameter Tuning and Best Results 47
5.4 Discussion and Conclusions 48

6 Continuously Participating vs. Certified Student Analysis **49**
6.1 Data Observations and Analysis 49
6.1.1 Activity vs. Grade Analysis 49
6.1.2 Delta Activity vs. Grade 51
6.1.3 Delta Activity vs. Delta Grade 55
6.2 Discussion 56
6.3 Conclusion 58

7 Principle Conclusions and Future Work **59**

List of Figures

3-1 Depiction of grade states(s_i), before activity, and after activity 29

4-1 Counts of certified students with grades within specific ranges at different dates for 6.00.1x Spring 2016. 32

4-2 Mean and standard deviation of before activity and after activity in the Spring 2016 offering of 6.00.1x 33

4-3 Spring 2016 6.00.2x delta activity vs. grade of all grade states. Line indicates no change in activity, $x = 0$. 36

4-4 Distributions of all offerings of 6.00.2x delta activity across all grade states 38

4-5 Fall 2016 Delta Activity vs. Delta Grade of All Grade States 40

4-6 Spring 2017 6.00.1x Delta Activity vs. Grade Separated Into Past Experience 43

4-7 Spring 2017 6.00.2x Delta Activity vs. Delta Grade Separated Into Past Experience 44

6-1 Depiction of the before and after activity vs. grade on certified and continuously participating students on problem set 2 of the Spring 2016 iteration of 6.00.1x. 50

6-2 Depiction of delta activity vs. grade on certified and continuously participating students on the Spring 2016 iteration of 6.00.2x. 51

6-3 Depiction of delta activity vs. delta grade on certified and continuously participating students on the Fall 2016 iterations of 6.00.1x and 6.00.2x. 52

List of Tables

1.1 Number of certified students and what percentage of total students they make up 17

3.1 Note that for 6.00.1x Fall 2016 our activity data set begins with the second problem set. Although there are 6 problem sets and 2 exams according to the course axis, our analysis is of 5 problem sets and 2 exams. 25

3.2 6.00.1x and 6.00.2x graded assessment type and topic 26

3.3 Past experience of certified students as indicated on their course entry survey. 26

4.1 P-Values from the normality test of all grade states of all offerings. White row is 6.00.1x and blue row is 6.00.2x. 34

4.2 Z-scores of skew tests on all grade states(assignment numbers) of all course offerings. White row is 6.00.1x and blue row is 6.00.2x. 35

5.1 This table shows the cross-validated accuracies of the best trained model of each type. 47

6.1 P-values of the kolmogorov-smirnov test (KS-Test) for testing that the delta activity distribution of certified continuously participating students is the same as the distribution of not certified continuously participating students. 52

6.2 P-values of the kolmogorov-smirnov test (KS-Test) for testing that the delta grade distribution of certified continuously participating students is the same as the distribution of not certified continuously participating students. 53

Chapter 1

Introduction

Massive open online courses (MOOCs) are a 21st century innovation in field of education. As MOOCs become more widespread, understanding their impact on students becomes more important. This research focuses on observing student behavior in MOOCs and what implications grades have on this behavior. These observations and analyses give us more insight on how students behave with the current course design and how MOOCs designs could be changed to better students in the future.

In this chapter, we will discuss MOOC design and lay the foundations for our research. We connect traditional learning models to our analysis of MOOCs. Throughout our analysis we use data from two EdX MOOCs: *6.00.1x Introduction to Computer Science and Programming Using Python* and *6.00.2x Introduction to Computational Thinking and Data Science*, offered in 2016 and 2017.

1.1 The Design of MOOCs

MOOCs are able to offer a "massive" amount of students access to education in a multitude of fields [9]. Course materials are available on the internet, offering students more flexibility in terms of their time and place of study. At the same time, many MOOCs can be taken with significantly lower financial cost than the same courses taught in a traditional school setting [3]. EdX, the MOOC platform created from a collaboration of MIT, Harvard, and other universities, is an example of a MOOC

exhibiting these qualities of high participation and low cost.

Though MOOCs of different platforms, and some within the same platform, have different designs, they generally try to emulate some aspect of traditional schooling, as these designs are a conventional way students currently learn. The EdX platform is a site that offers a wide variety of MOOCs that use these strategies to create online lessons. Course material is split into units, and each unit is equipped with online lecture videos, lecture notes, and a homework problem set. At the middle and end of the course, there is a time-limited assessment that test the students' overall knowledge of course material. Using time-limited releases of problem sets and assessments, MOOCs on the EdX platform can emulate the traditional education lesson plans. Students are led to learn one unit at a time in a purposeful manner, and feedback through grades on course assignments allow students to see how well they are keeping up with the course material and what their final grades will be. Finally students can receive a certification at the end of the course if their overall grade is above a specific threshold and they pay a small fee for certification.

1.1.1 Certification in MOOCs

On the EdX and other MOOC platforms, students are given a certification when they meet the requirements of passing a course. Many students take MOOCs to solely get this certification. MOOC certification can be used as a validation for certain skills taught in the course. For example, a certification in a basic programming course could be some proof of ability in basic programming. Although, these certifications cannot be likened to degrees, they can still be useful in various fields.

Due to the high availability and scalability of some MOOCs, registration can be on the order of thousands (such as the MOOCs we are observing). However, though the interest and capacity to teach students are high, the actual participation and completion of the course by registered students is low. It is estimated that only 10 percent of students who register for MOOCs actually do all of the assignments to completion [16][8]. Table 1.1.1 shows the number of certified students in relation to the number of total students in MOOCs that we observe for this study.

Number of Certified vs. Total Students

Course Number and Offering	Number of Certified Students	Number of Total Registered Students
6.00.1x Spring 2016	947 (2.3%)	40749
6.00.2x Spring 2016	518 (2.4%)	22041
6.00.1x Spring 2017	1511 (2.2%)	69420
6.00.2x Spring 2017	324 (1.8%)	18290

Table 1.1: Number of certified students and what percentage of total students they make up

1.1.2 Grades, Dropout, and Disengagement

One of the possible factors in whether an individual student participates in the MOOC to completion is their performance (in terms of how well the student does on graded assignments). This is addressed as one of the underlying factors of disengagement in the paper Kizilcec et al.[12]. To better understand how performance can impact the decisions they make to participate or disengage from the MOOC, we analyze the relationship between student activity and the grades students receive.

In order to study specifically disengagement and grades, we look into students who do all assignments up until a specific assignment and how they differ from (or are the same as) certified students. Students who do assignments (up to a certain point) and certified students have a high level of motivation. Therefore, similarities and differences between the two will be less influenced by fluctuations in motivation.

1.2 Self-Regulated Learning Systems and Feedback

Though MOOCs' designs tend to imitate conventional classroom learning models, they require a lot more self-regulation to do well and succeed in learning course materials. Students are able to access class materials at any time and can do assignments at any time within the set time limitation. Because they do not receive as much personal instruction as in a traditional classroom, grades are an important mode of feedback for students. Feedback is widely observed to influence student behavior,

especially in self-regulated learning systems [1]. For example, a student who receives a poor grade would hopefully become more motivated to do better and change his or her learning behavior whereas a student who receives a good grade would reinforce his or her current learning methods. Thus, understanding of the impact of grades on student learning behavior is useful for course designers in anticipating student behavior in current and future courses.

Self-regulated learning models have been widely theorized and debated [2]. In the following sections, we will connect those learning models to our analysis of student activity in MOOCs.

1.3 Research Questions

In this thesis, we will discuss the hypothesis and results of 3 separate mini-studies on the relationship between student activity in MOOCs and grades. These studies will be motivated by the following research questions:

1. How do certified students change their behavior given the grade they receive? How does past experience effect this?

2. Can we find a non-linear relationship with machine learning models? Specifically can we predict grade with activity?

3. From our observations on certified students, do we see a similar pattern in students who continuously participate in MOOC assignments? Do grades effect dropout from continuously participating students?

The first study will focus on the analysis of the impacts of grades on student behavior in certified students only. We will then extend upon the first study and focus on relating to past knowledge to student activity and grades. The second study will focus on machine learning models applied activity and past experience and how these models can be used to predict grade. The third study will focus on a contrast between the results of the first study on certified students and students who continuously participate up to certain assignments in MOOCs.

Chapter 2

Background and Related Works

In this section, we look at previous work in this field of MOOCs and the modeling of learning behavior. We will discuss the implications of theoretical models and how they are used to formulate our hypothesis on student learning behavior. We will also discuss previous experiments and their results on the impact of grades in MOOC students and how they shape our hypothesis on grade impact.

2.1 Performance Prediction

Individual performance prediction has been widely studied to understand what factors have significant impact on student grades in MOOCs. Motivation and participation in assessments and forums are shown to be a key indicator of performance [4]. Student activity, in the form of accesses to course material is shown to be correlated to student performance [17]. When working with time series data, Ren et al.[17] have shown that previous grades are a significant indicator of future grades. Jiang et al. [10] show in their study that the grades received in the first week (the first assessment grade) is a good predictor of final grade.

In our second study (Chapter 5) we will discuss our own models for predicting grade increases and decreased in MOOC students. We will also discuss the factors we found most significant in our analysis and how well they predict grade trajectory.

2.1.1 Drop Out Prediction

Due to the overwhelming dropout rate in MOOCs, many studies have been performed to better understand why students "drop out" or become classified as not certified participants. In a study by Kizilcec et el. [12], not certified participants can be further split into 3 behavioral categories: auditing students, disengaging students, and sampling students. Auditing students are those that browse through the course material but do not take assessments and drop out; disengaging students are those that begin the course doing all assessments but stop participating and drop out part-way through; and sampling students participate in few (usually only 1) chapter or assessment period in their entirety as a registered student. [12]. Though different studies may reference these behaviors with different names, this categorization creates a clear distinction between the behaviors of subsets of not certified MOOC participants. Furthermore, the separation of the not certified student into behavioral subsets can give a better explanation as to why students drop out and if the dropout could be prevented [19].

When trying to predict individual and total class dropout, grades, activity, and class characteristics have shown to be highly correlated to drop out [11] [8]. In a study by Halawa et al.[11] nearly all students who received 0.5 or less (points earned divided by points total), eventually dropped out of the course. One possible reason for this is that the low score indicates to the student that they will not be successful in the course, and therefore quits. Another explanation could be that students with continuously low scores can no longer meet the grade cutoff to pass. Activity, studied at both high and low granularity, has shown to be indicative of dropout [21] [13]. The reason for this is that higher activity and interaction with peers shows higher engagement and motivation, which is negatively correlated to dropout [4][20]. Finally, the timing and design of the course also greatly influences the collective dropout rate [11]. In exit surveys, time pressures to look over course material and do assignments is often cited as one of the most important reasons to for dropout.

After the identification of relevant features, MOOC data can be used to build

relevant predictive models. Due to the multitude of possible features, neural networks are often used as the base model in dropout prediction [18][5]. Fei et al. [7] used a recurrent neural network to capture time series data relationships. However, a higher complexity model is not a better predictor of MOOC participant dropout. White et al. [18] researched the performance of higher layered neural networks and found that they under-perform simpler MOOC dropout models. Even so, MOOC dropout analysis and prevention is a complex problem whose study can help further improve the MOOC learning system.

Although we do not directly analyze dropout, in our third study (Chapter 6) we will be looking at the behavior of continuously participating students when they drop out and the factors that may contribute to this behavior.

2.2 Self-Regulated Learning Feedback Systems

MOOCs operate with a high level of self-regulated learning on the part of the student. Course designers study the impact of feedback on student learning behavior to achieve successful course designs [1]. Researchers in the field created models that hypothesize how students change their learning behavior after receiving feedback. Carver and Scheier developed a model that states that students will evaluate their feedback and adjust their learning behavior to reach their course goals [2]. For example, if a student has a goal of passing a course, upon receiving a lower grade on an assignment, the student will work harder to get better future grades to meet the goal of passing the course. Magill argues that this model inconsistent with different feedback mechanisms; Magill claims that feedback in the form of physical interactions provide different results than grade feedback [14]. Other researchers have argued that misinformation and inaccurate perceptions of goals contribute more to behavior changes than the original Carver and Scheier model[1]. Still, when tested in a traditional classroom, Main and Ost found results consistent with the Carver and Scheier model [15]. Following this work in the field, we will connect these learning models to student behavior in MOOCs.

Grades and other performance metrics have been studied in MOOCs. Ren et al. show that previous grades are a significant indicator of future grades and the students' final grades [17] . In Halawa et. al's [8], it was shown that grades are related to student dropout. Students with total grades of 50 percent or less eventually dropped out of the course [8]. One possible explanation for this is that the low grade is indicative that the student will not pass the course, therefore the student will quit instead of changing behavior. In chapter 6, we will compare the differences in grades in certified students and students who complete all assignments, discussing how grade may impact course completion and certification.

Chapter 3

Data Set and Research Methods

In this chapter, we will provide an overview of our data set and its properties. We will also define our research questions and the terminology we use throughout the rest of the experiments and data analysis.

3.1 Certified and Continuously Participating Students Definition

As previously explained, certified students are students who obtain a passing grade at the end of the course and pay a fee for certification. We will refer to *certified students* as students who have completed both criteria at the end of the course.

Continuously participating students are defined as students who continuously participate in all assignments up to a certain assignment. As shown in Table 3.2, continuously participating students and certified students overlap in their population. Continuously participating, at the end of the course, are students who attempted all assignments but did not receive a certification due to low grades or unwillingness to pay the certification fee.

3.2 Data Set

For this thesis, we use click stream and platform data from students from three offerings of the two EdX MOOCs *6.00.1x Introduction to Computer Science and Programming Using Python* and *6.00.2x Introduction to Computational Thinking and Data Science* in 2016 and 2017. Both 6.00.1x and 6.00.2x are courses that teach computer science concepts and programming. 6.00.2x is the sequel course to 6.00.1x and covers more advanced topics on the assumption that the student has knowledge of 6.00.1x topics. 6.00.1x and 6.00.2x do not offer collaborative learning, self assessment, or peer assessment. All grades are given by the courses' automatic grading system.

Each course is composed of multiple units which have an associated graded problem set. Each unit also has multiple lecture videos, notes, and optional "finger exercises" which are problems given to students that can be checked by the automatic grading system but do not factor into students' final grade. For our study, we will be using click stream data generated from these sources. Click stream data consist of all events a student can trigger on the site, including playing lecture videos, accessing unit notes, doing finger exercises, and completing problem sets or assessment tests. Click events are in total in the order of millions, with each student triggering hundreds to thousands every week.

Table 3.2 shows the breakdown of certified students and continuously participating students at the end of the course. It also shows the percentage each group in relation to all students who take the course. As shown in table 3.2, students who do all assignments are not necessarily certified (as they may not have a high enough grade or did not pay the certification fee), and students who are certified do not necessarily do all assignments (because they do not need to do all assignments if they are able to get enough points from the other assignments *and* because the course will drop the lowest graded or not attempted homework assignment). Therefore, observing the changes in activity of continuously participating and certified students captures the behavior of students who complete the course versus and those who eventually drop out. For a more consistent comparison, we use the raw points the students obtain

Number Certified vs. Continuously Participating Students

Course	Offering	Number of Certified Students	Number of Continuously Participating Students	Number in Both	Number of Assessments
6.00.1x	Spring 2016 (s2016)	947 (2.3%)	1,021 (2.5%)	610 (1.5%)	7
6.00.2x	Spring 2016 (s2016)	518 (2.4%)	572 (2.6%)	303 (1.4%)	6
6.00.1x	Fall 2016 (f2016)	2,086 (3.3%)	2,468 (3.9%)	1245 (2.0%)	8
6.00.2x	Fall 2016 (f2016)	528 (2.9%)	643 (3.5%)	350 (1.9%)	6
6.00.1x	Spring 2017 (s2017)	1,511 (2.2%)	1,638 (2.4%)	882 (1.3%)	8
6.00.2x	Spring 2017 (s2017)	324 (1.8%)	456 (2.5%)	227 (1.2$)	6

Table 3.1: Population Breakdown of Certified and Continuously Participating Students [2]

divided by total possible points in attempted assignments in the grade analysis of the problem sets and assessments.

3.2.1 Course Content

In Table 3.2.1, we detail the types of assessment (Problem Set or Quiz) as well as their general content. The exact number of problem sets was different for some offerings of 6.00.1x, but the chapter content stayed the same.

3.2.2 Past Experience Data

In addition to click stream data, we will also use survey data collected at the beginning of the course about the students' prior experience of the course material. Because we are analyzing programming courses, this includes past programming experience and known languages. Additionally, because 6.00.2x is an extension the

6.00.1x and 6.00.2x Syllabus

Topic	6.00.1x Assignment #	Assignment Type	Topic	6.00.2x Assignment #	Assignment Type
Python Basics	1	Problem Set	Optimization	1	Problem Set
Simple Programs	2	Problem Set	Randomness	2	Problem Set
Midterm	3	Exam	Midterm	3	Exam
Structured Types	4	Problem Set	Statistics	4	Problem Set
Good Programming Practices	5	Problem Set	Modeling and Fit	5	Problem Set
Object Oriented Programming	6	Problem Set	Final	6	Exam
Algorithmic Complexity	7	Problem Set			
Final	8	Exam			

Table 3.2: 6.00.1x and 6.00.2x graded assessment type and topic

Previous Experience of Certified Students

Course Number and Offering	No Experience	Know Python	Other Language	No Response	Veteran	Took 001X
6.00.1x Fall 2016	452	316	951	206	46	N/A
6.00.2x Fall 2016	1	39	N/A	209	N/A	279
6.00.1x Spring 2017	401	261	617	203	29	N/A
6.00.2x Spring 2017	1	27	N/A	151	N/A	143

Table 3.3: Past experience of certified students as indicated on their course entry survey.

concepts from 6.00.1x, 6.00.2x students also indicate whether they have taken 6.00.1x.

Table 3.2.2 shows the breakdown of the past experiences for certified students. "N/A" values are for categories that did not exist on that offering's survey. As shown in Table 3.2.2, the past experiences of certified students is varied, and the majority have some experience in coding. This is not surprising as these are also the students who eventually pass the course and receive certification.

3.3 Overview of Analysis

We will now address the principle questions of our research, our hypothesis regarding each of those questions, how we plan to test those hypothesis. Once again, we intend to answer the following questions:

1. How do certified students change their behavior given the grade they receive? How does past experience effect this?

2. Can we find a non-linear relationship with machine learning models? Specifically can we predict grade with activity?

3. From our observations on certified students, do we see a similar pattern in students who continuously participate in MOOC assignments? Do grades effect dropout from continuously participating students?

3.3.1 Certified Students Analysis

Grades can be derived to convey how well a student comprehends a subject matter and how well students apply their comprehension. Grades also determine whether a student passes, or is allowed to become certified in, a specific course. This certification is valuable to students as it can be used to demonstrate the student has certain skills or qualifications. To obtain certification, we hypothesize that certified students will noticeably adjust their behavior to the grades that they receive. Specifically, our initial belief is that students with lower grades will raise their activity levels to improve their grades and students with high grades will maintain their activity and grades in order to obtain their certification.

In a review of our research with the course designers of 6.00.1x and 6.00.2x, it was suggested to explore our results in relation to the past experience of the students. Therefore we re-apply our analysis to separate students of different experience levels and observe the results.

3.3.2 Grade Trajectory Analysis

The complexity of the relationship between grade and activity is not necessarily linear. Machine learning models are used in the analysis and prediction of multi-variable problems with non-linear relationships. As low grades have been researched to be associated with dropout [8], it would be valuable to learning designers to understand which factors correlate to when students' grades decrease. Therefore we apply three machine learning models– logistic regression, classification tree, and neural networks–on activity and past experience variables to ultimately predict whether students' grades drop. We will tune these models' hyper-parameters to increase their performance and comment on what this performance means for the relationship between grades, activity, and prior experience.

3.3.3 Continuously Participating vs. Certified Students

In accordance with the learning model by Carver and Scheier[2], we hypothesize that students will change their learning behavior after receiving grade feedback. Specifically, we believe that certified students will adjust their behavior to pass the course (students with lower grades would work harder and students with higher grades would maintain them with steady activity levels) and that continuously participating students would exhibit similar but less pronounced behavior changes as certified students (as some are not motivated by certification to complete the course). We also hypothesize that continuously participating students with lower raw grades (less than 0.5) will drop out or not receive certification, as described by Halawa et. al [8].

3.4 Key Terms and Definitions for Analysis

We will now define the terms relating to grades and student activity that we will use throughout our analysis.

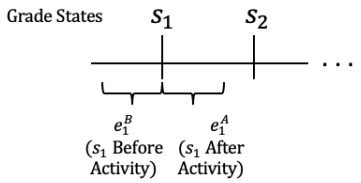

Figure 3-1: Depiction of grade states(s_i), before activity, and after activity

3.4.1 Grade States

We define the *grade state* of a student as the cumulative grade (total points earned divided by total possible attempted points earned) at the time when a grade is finalized. All grades are finalized and released at 23:30:00 UTC on the due date of each problem set assignment and exam assessment. In the context of our *continuously participating* student definition, continuously participating students at grade state s_i are all students who have attempted all assignments up to and including assignment i. That is, any continuously participating student with a state s_i also has states $s_1, ...s_{i-1}$, where s_1 is the grade state of the first assignment.

3.4.2 Cumulative Activity

Next, we define the *cumulative activity* e_i of a student as the total number of click events the student triggers in a 7-day period. Although we use a 7-day period of observation, grade states are not necessarily 7 days apart. Therefore the $e_i^A \neq e_{i+1}^B$. Next, we define the *before activity* e_i^B as the cumulative activity before a grade state is finalized, and we define the *after activity* e_i^A as the cumulative activity after a grade state is finalize. Figure 3-1 shows the relationship between before and after activities at a specific grade state.

3.4.3 Delta Activity and Delta Grade

We lastly define the *delta activity* Δe and *delta grade* Δs. *Delta activity* is the difference between the after and before activity of a student at a specific grade state. *Delta grade* is the change in cumulative grade from grade state s_i to grade state s_{i+1}.

Therefore we define the equations:

$$\Delta s_i = s_{i+1} - s_i$$
$$\Delta e_i = e_i^A - e_i^B$$

Because delta activity is the difference of before and after activities, it exists for all assignments (although it may be skewed due to when the assignment occurs within the course). Meanwhile, delta grade only exists after an initial grade is given (ie. it exists for grade states s_2 and beyond).

Chapter 4

Certified Student Analysis

We will now discuss our first study on examining the behavior of certified students our MOOCs. Specifically we hypothesize that students will noticeably change their behavior as they receive certain grades that push them towards their goal of getting a certification.

4.0.1 Grade Breakdown

To better understand the observational MOOC data sets, we break down the grades for each offering of each unit. For all offerings of all units, most students predominantly received grades higher than 90 percent, while very few students (if any) pass the course with a cumulative grade below 70 percent. Although a student only needs to finish the course with a grade higher than 65 percent to pass, we found no instances where a certified student ever dropped below 60 percent cumulative grade.

We calculate the cumulative grade at each grade state by summing the total points earned at the time of the new grade release and divide by the total points possible in the course at that time. Figure 4-1 shows the breakdown by assignment of how many students receive what grade in the Spring 2016 offering of 6.00.1x. The number of students with cumulative grades higher than 80 percent drops as the course progresses.

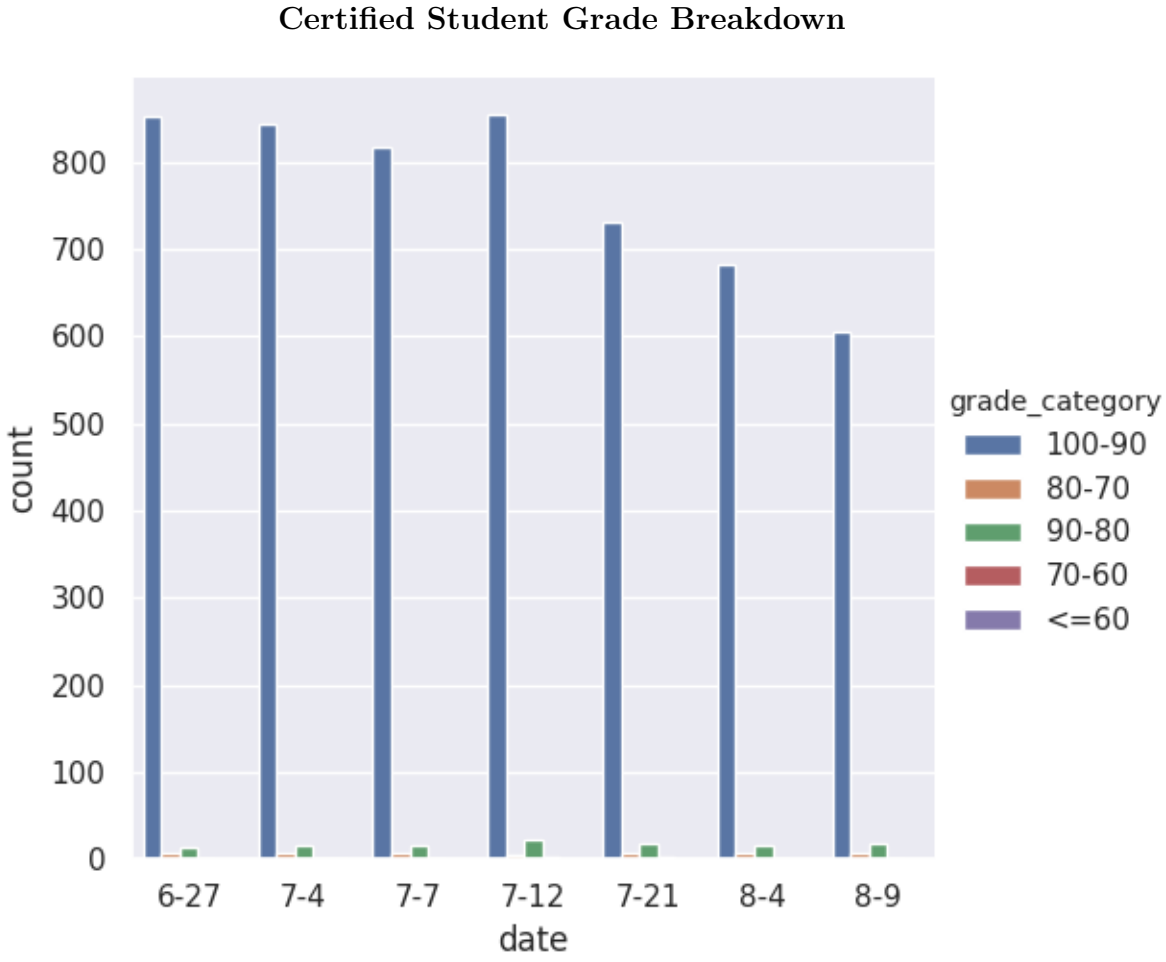

Figure 4-1: Counts of certified students with grades within specific ranges at different dates for 6.00.1x Spring 2016.

4.0.2 Activity Analysis

Similarly, we analyze the before activity and after activity of students to better understand how much activity students typically have and how much variation there is among students. All course offerings observed similar mean and standard deviation figures with most student activity being between 0 and 2,000 click events triggered.

For each grade state of each course offering, we calculate the before and after activity by summing the total daily click events triggered for the 7 days before and after the next grade state. Because we are looking at the change in behavior after transitioning to a new grade state, we keep the activity calculation conditions the same for all grade states.

Figure 4-2 shows a time plot of the distributions of before and after activity of the Spring 2016 offering of 6.00.1x. The x-axis shows the date of the grade state and the y-axis shows the cumulative activity. The before and after activity means stay close together in the beginning and middle of the course hovering at or above 500 click events. Afterwards, after activity consistently falls towards the end of the course. Finally there is a large gap between the last before activity and after activity points. This is due to the increase of activity to study for finals and the lack of necessary activity once the final assessment score is given and the course is over.

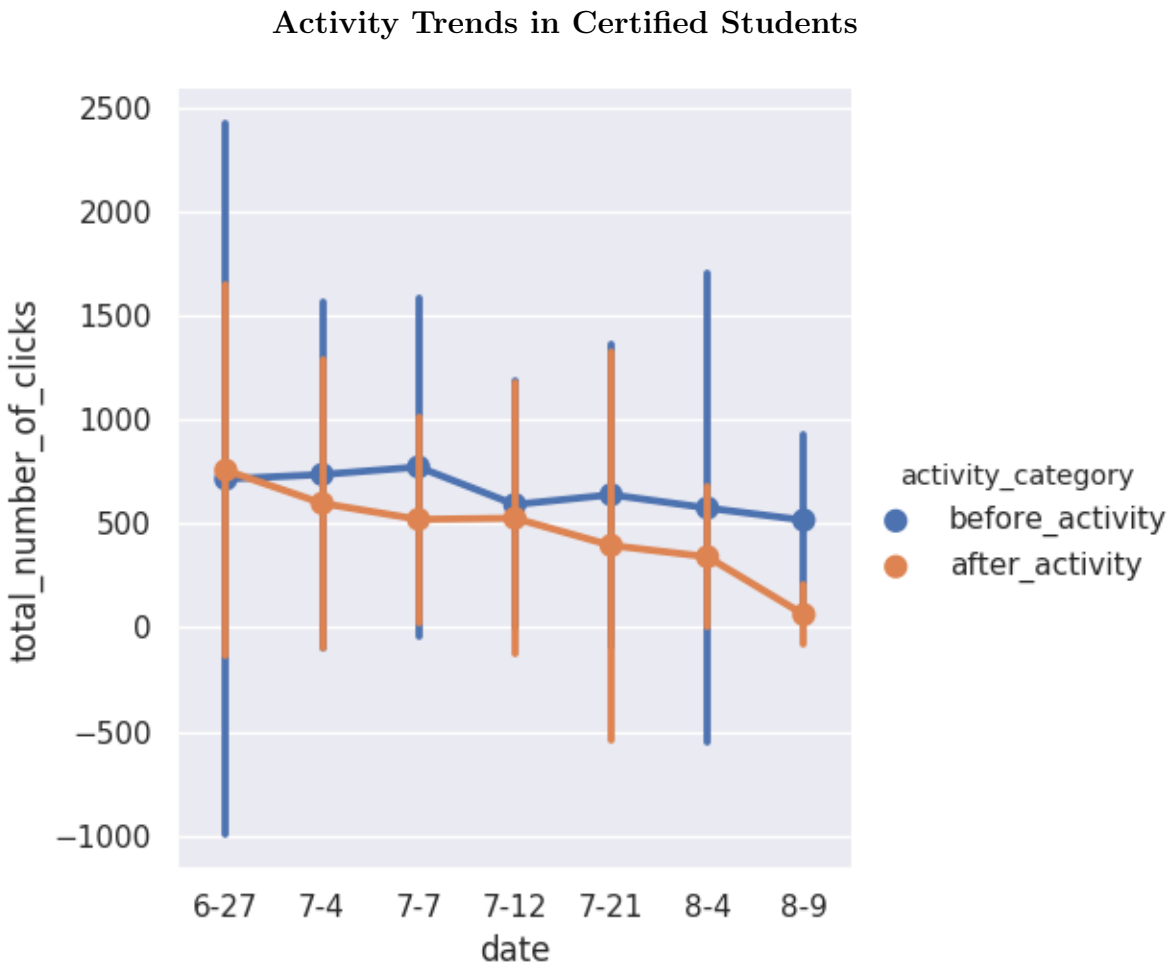

Figure 4-2: Mean and standard deviation of before activity and after activity in the Spring 2016 offering of 6.00.1x

Delta Activity P-Values

	Assignment Number							
Course	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
6.00.1x f2016	6.0648 e-117	0.0	7.2073 e-90	0.0	4.3680 e-154	0.0	0.0	
6.00.2x f2016	2.5519 e-64	1.8363 e-121	0.9131	0.1804	3.5404 e-3	6.5653 e-22		
6.00.1x s2016	4.1068 e-3	0.0	7.2471 e-20	8.0116 e-190	5.9559 e-163	2.1309 e-284	5.1281 e-270	1.4095 e-23
6.00.2x s2016	6.5224 e-3	0.6470	1.1900 e-2	1.2952 e-2	2.6773 e-3	1.8578 e-14		
6.00.1x s2017	0.0	0.0	0.0 0.0	0.0	0.0	0.0	0.0	
6.00.2x s2017	3.3758 e-90	2.7355 e-5	0.7782	3.0856 e-5	2.1081 e-2	4.4132 e-14		

Table 4.1: P-Values from the normality test of all grade states of all offerings. White row is 6.00.1x and blue row is 6.00.2x.

4.1 Analysis on Delta Activity

Following our initial data analysis, we focus our study on how behavior before and after a grade is released compares to the grade received. After we find all before and after activity for all grade states, we calculate the delta activity by taking the difference between the after activity and the before activity. For most grade states, we find distribution of delta activity is centered at or below 0, indicating that on average, students do less activity as the course progresses. We perform a normality test for each of the delta activity distributions with null hypothesis: the sample is from a normal distribution. Table 4.1 shows the p-values of that this null hypothesis is true using the `scipy.stats.normaltest` function[1]. The p-values indicate that delta activity might not be normally distributed.

We furthermore find the skew of each delta activity data set using the `scipy.stats.skewtest` function[2]. We calculate the z-score of the null hypothesis that the distribution has the same skew as a normal distribution, shown in Table 4.1. We observe the that magnitude of the delta activity skews of 6.00.1x are systematically larger than those

[1]https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.normaltest.html
[2]https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.skewtest.html

Delta Activity Z-Scores

	Assignment Number							
Course	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
6.00.1x f2016	-8.006	-1.605	-3.836	2.700	7.865	-14.758	-0.631	
6.00.2x f2016	1.977	-20.997	-1.577	16.328	-2.658	-6.747	-11.574	
6.00.1x s2016	1.766	-16.306	-8.073	-7.642	5.777	16.479	-15.741	-1.115
6.00.2x s2016	0.302	0.050	0.278	-0.293	-0.347	-1.224		
6.00.1x s2017	30.814	37.350	-37.707	-37.427	-37.162	33.803	-34.694	-29.507
6.00.2x s2017	11.705	0.615	0.041	-0.655	-0.328	-1.423		

Table 4.2: Z-scores of skew tests on all grade states(assignment numbers) of all course offerings. White row is 6.00.1x and blue row is 6.00.2x.

of 6.00.2x.

For each grade state of each offering, we plot the delta activity against the new cumulative grade. Figure 4-3 shows the delta activity vs. cumulative grade scatter plots for each grade state of the Spring 2016 offering of 6.00.1x. We find that variation about the mean of the distribution decreases as grade decreases. The largest magnitudes of delta activity are found in students with the highest grades. Meanwhile, the individuals with the lowest grades mostly have delta activity close to 0 or negative delta activity.

As a whole, most scatter plots of delta activity vs. cumulative grade seem to mirror the distribution of delta activity. This can be shown in Figure 4-3, by the top horizontal histogram and center scatter plot.

4.1.1 Qualitative Analysis of Delta Activity vs. Grade

After conducting our initial delta activity analysis, we connect our data results with implications on observed student behaviors. Concretely, points on the the delta activity vs. grade scatter plot can be split into two sections: points to the left of the $x = 0$ line (the "left" section) and points to the right of the $x = 0$ line (the

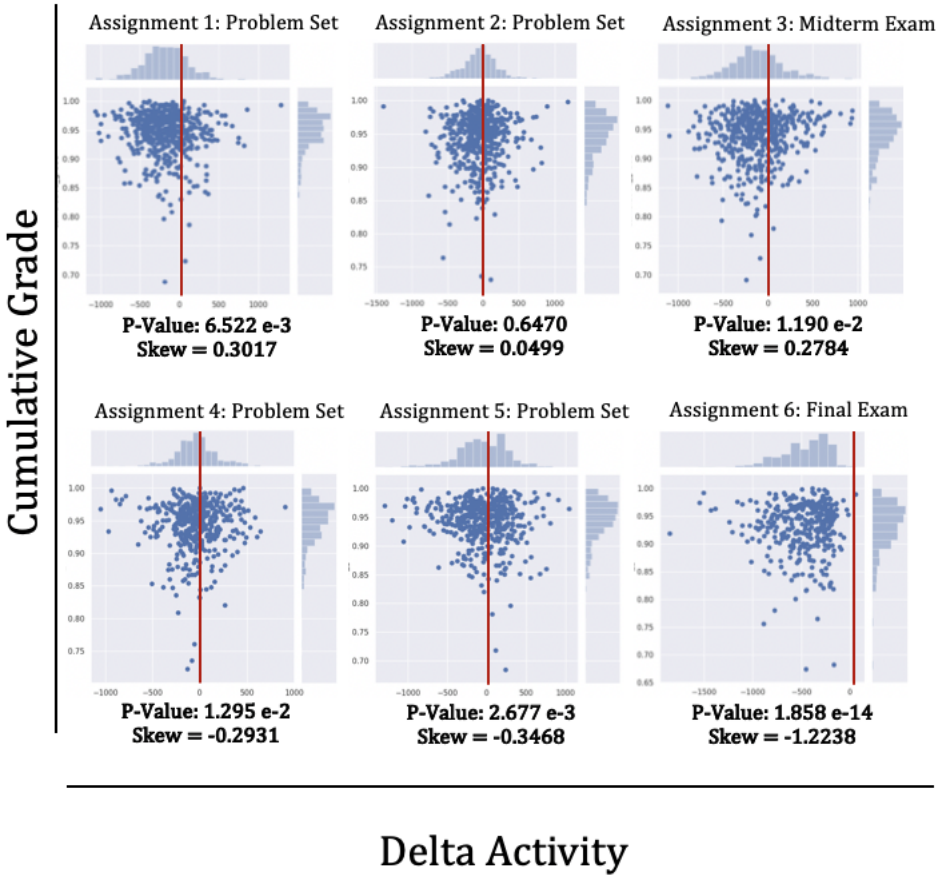

Figure 4-3: Spring 2016 6.00.2x delta activity vs. grade of all grade states. Line indicates no change in activity, $x = 0$.

"right" section). The significance of the left section is that it shows the students who became less active after receiving their grade and what grade caused them to do so. Similarly, the right section shows the students who became more active after receiving their grade and the grade that they received.

Relatively, we can define 4 extremes of changing behavior as the 4 corners on the delta activity vs. grade graphs. The upper-left corner shows students who decreased their activity drastically after receiving a high grade. The upper-right corner shows students who increased their activity drastically after receiving a high grade. The lower-left shows students who decreased their activity after receiving a low grade. Finally, the lower-right shows the students who increased their activity after receiving a low grade.

In Figure 4-3, we can see a general trend of decreasing activity as the course progresses, indicated by the more populated left sections of the delta activity vs. grade plots. We can also see many students doing drastically more or drastically less activity after receiving a high grade. However, we do not see many instances increasing or decreasing activity after receiving a lower grade. The latter can be explained by how we selected our sample: we used only certified students that completed the course, whereas students who decreased their activity while receiving poor grades would have likely dropped out. The former observation suggests that students who receive poor grades (given that they do not drop out) also do not raise their activity level. In fact, the lowest scoring students, in general, have close to 0 change in their activity after receiving their grade.

4.2 Comparisons Across Courses and Course Offerings

We now shift from individual course analysis to how distributions compare across courses and offerings. We investigate how changes in cumulative grade correlate to changes in activity and how these changes differ by course.

4.2.1 Delta Activity Across Offerings

Following our delta activity analysis on individual course offerings, we compare delta activity distributions across all offerings of the same course. Figure 4-4, shows a boxplot of the delta activity across all assignments compares across all offerings of 6.00.2x. The center point describes the median delta activity of the assignment, while the wings indicate variability outside the upper and lower quartiles of the data set. Figure 4-4 illustrates that the means of delta activity for all assignments across all offerings stay at around or below 0, with the exception of the last assignment. Though there is no discernible pattern across offerings, the final assessment in all offerings shows general decrease in activity as the class ends.

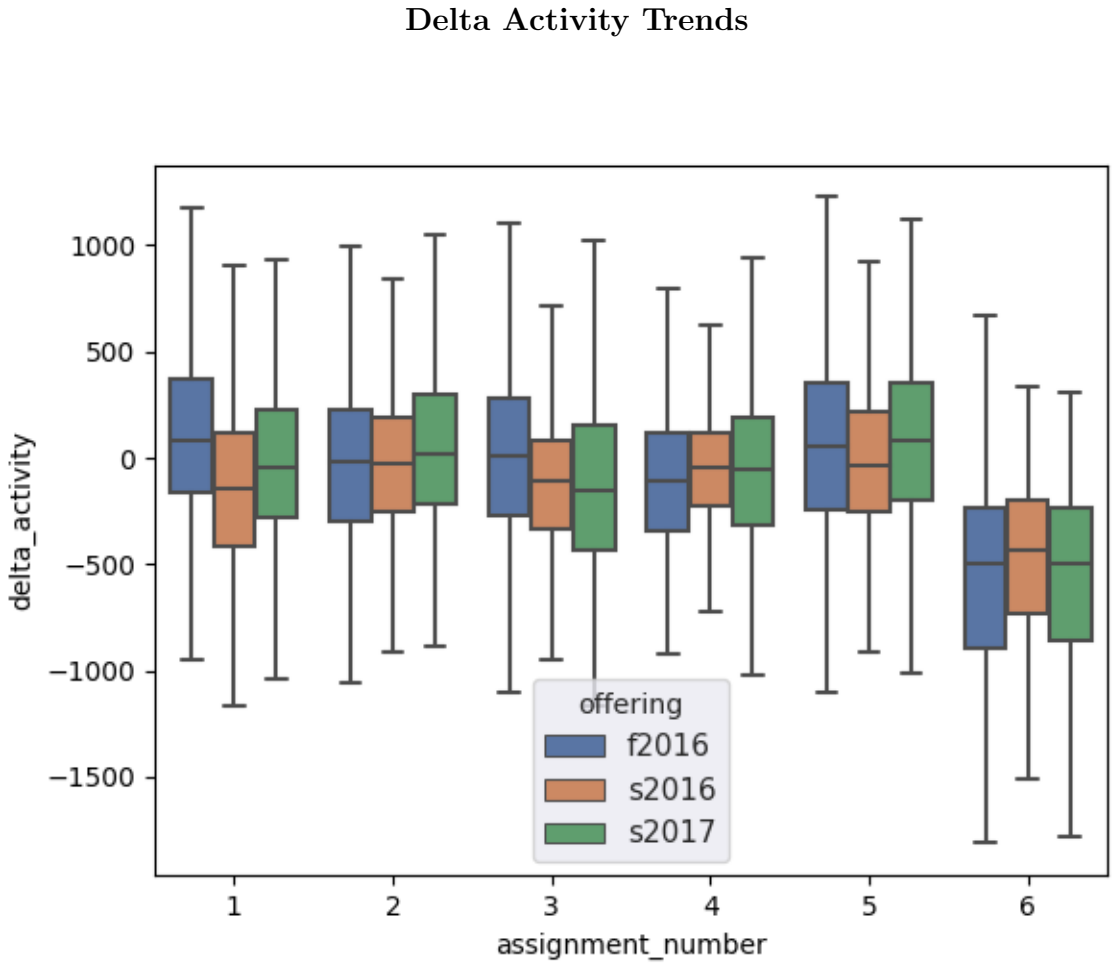

Figure 4-4: Distributions of all offerings of 6.00.2x delta activity across all grade states

Because 6.00.1x does not have the same number of assessments per offering, we are unable to align and compare them in the same way. However, the 6.00.1x offerings individually show a the same trend (or lack thereof) as the 6.00.2x offerings.

4.2.2 Delta Activity vs. Delta Grade

To further understand the implications of grade on activity, we investigate the how delta activity correlates to changes in grade (delta grade). We calculate the delta grade by taking the difference between the grade at some grade state s_t and the grade at s_{t-1}. Figure 4-5 shows delta activity vs. delta grade scatter plots, for all assignments in the Fall 2016 offering of 6.00.1x and Fall 2016 offering of 6.00.2x.

Qualitative Significance

Using our plots of delta activity vs. delta grade, we can observe more implications about student behaviors. We can divide the delta activity vs. delta grade graphs into four sections along $x = 0$ and $y = 0$. The upper-left section shows students who raised their overall grade while having less activity. The upper-right sections shows students who raised their overall grade while having more activity. The lower-left shows students who dropped in grades while having less activity. Lastly, lower-right shows students who dropped in grades despite increasing their activity.

The 11 scatter plots in Figure 4-5, show that all 4 extremes of the 4 sections are uncommon and rarely occur. Moreover, the largest changes on one axis are often along the 0 line of the other axis (we will refer to this as the "star" formation). This observation then supports the claims that the largest variations in activity occur for students with little or no change in grade and while the largest variations in grade occur when students have little or no change in activity.

Differences Between 6.00.1x and 6.00.2x

The major difference between 6.00.1x and 6.00.2x data sets occur with the delta activity vs. delta grade graphs. As seen in the Fall 2016 offerings in Figure 4-5,

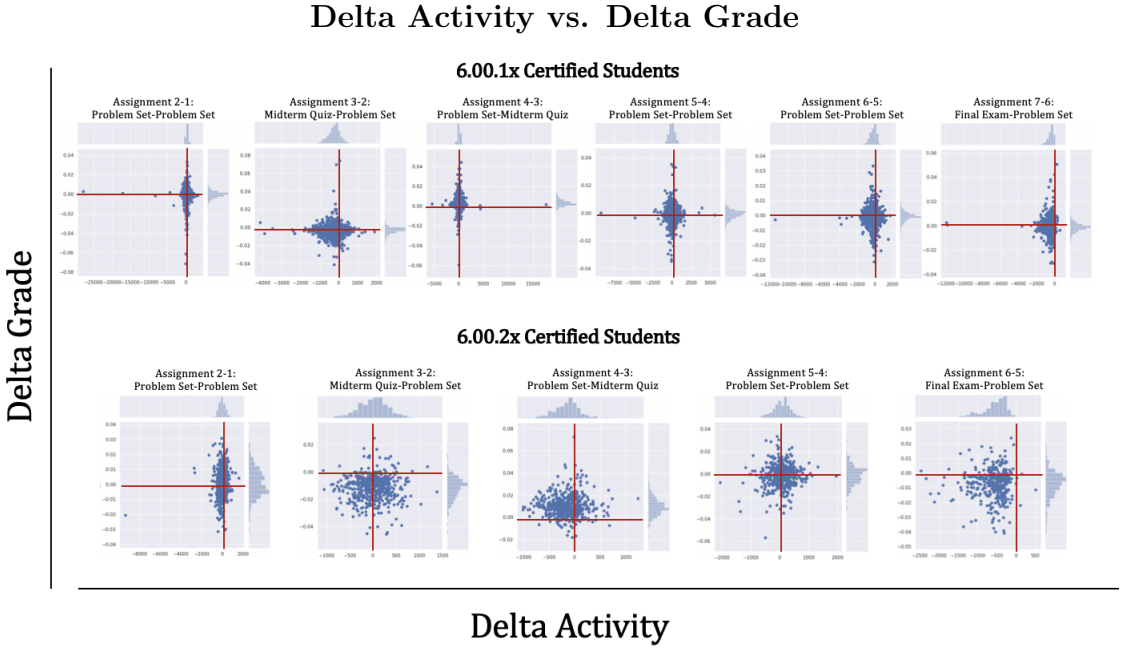

Figure 4-5: Fall 2016 Delta Activity vs. Delta Grade of All Grade States

6.00.1x has very defined "star" shapes for nearly all delta activity vs. delta grade graphs. This means that for every change in grade, the largest grade change occurred for students who minimally changed in activity, while the largest changes in activity occurred for students who had minimal changes in grade.

Though this "star" shape is also apparent in some 6.00.2x plots, we notice that the shape is not as defined, and mostly not centered at x=0, y=0. This trait is present in all three offerings of 6.00.2x, making it more likely that these traits are systematic rather than coincidental.

4.3 Discussion

Through the course of our investigation, we were focused on finding and studying data correlations between student grade and activity and tie this to student behavior and course design. Our initial belief was that students who have lower grades will try to raise them and students with high grades would maintain their grades and activity. However, our observations do not support these beliefs and instead suggest that these

behaviors do not happen at all in our MOOC data sets.

4.3.1 Initial Analysis

In our initial analysis of grade and activity, we find a decrease of grades and activity as the course goes by. We can speculate that this is due to the increase in difficulty of concepts as the course progresses and fatigue from the students. Additionally, we observe a decrease in activity at the end of all courses, which could be due to the course ending or students higher grades putting in less effort because they do not need to. These observations were expected and consistent with course design.

4.3.2 Delta Activity

The delta activity analysis supports the claims that those with high grades have higher variation in activity while those with lower grades have lower variation in activity. This is the opposite of our initial hypothesis and introduces new questions on why this occurs instead of the contrary.

One possible explanation for the increase in delta activity variation among high performing students is that these students already have a high understanding of the subject and do assessments at varied times because they do not need to continuously learn new material. This opens future studies into stratifying students by past experience and how past experience influences behavior.

A possible explanation for the small variation in delta activity for students with lower grades is that these students do not want to or do not have time or ability to improve their grades. Students maintain their activity habits to maintain their non-failing grades and pass the course. These individuals could also be changing their behavior in the type of course materials they access instead of how many. This opens possible future works into looking at the specific work activities students do on the site, instead of the total click events triggered.

4.3.3 Delta Grade

Delta grade shows how student performance changes over the progression of the course and how this could be related to changes in activity. However, our results do not indicate there is universal correlation between the two. Instead we find that the correlation to changes in grade and activity can be dependent on the course being studied and how it is designed. For example 6.00.1x is a purely introductory course that does not assume prior knowledge. In this instance, it is shown that higher grade increases or decreases correlate to lower changes activity and vice versa. From a design standpoint, this could only make sense if confounding factors were in play. On the other hand, 6.00.2x has little or no patterns in its delta activity vs. delta grade plots across all grade states, indicating the confounding factors in 6.00.1x may not be the same or present in 6.00.2x.

Another possibility is that delta activity vs. delta grade is related to the inherent "difficulty" of the assessment with respect to the students' base knowledge and comprehension of course materials. We find that the shape of the delta activity vs. delta grade least resembles a "star" shape centered at (0,0) for when the assessment is an exam (a midterm or final). Furthermore, because 6.00.2x assumes knowledge from 6.00.1x, the course as a whole would be "more difficult" than 6.00.1x, resulting in fewer "star" shapes and distribution centers below x=0. This hypothesis is consistent with our findings and would be viable area of future work.

4.4 Past Experience

Figures 4-6 and 4-7 show distributions of delta activity vs. grade and delta activity vs. delta grade of the Spring 2017 offering of 6.00.2x. Throughout the graphs, people of various past experiences are scattered with no particular pattern. In fact, students who have prior knowledge in python or some other programming language do not out perform students who have no prior coding experience.

One reason for this could be the bias of the data set of certified students. Because certified students all excelled enough in the course to pass, these students are not

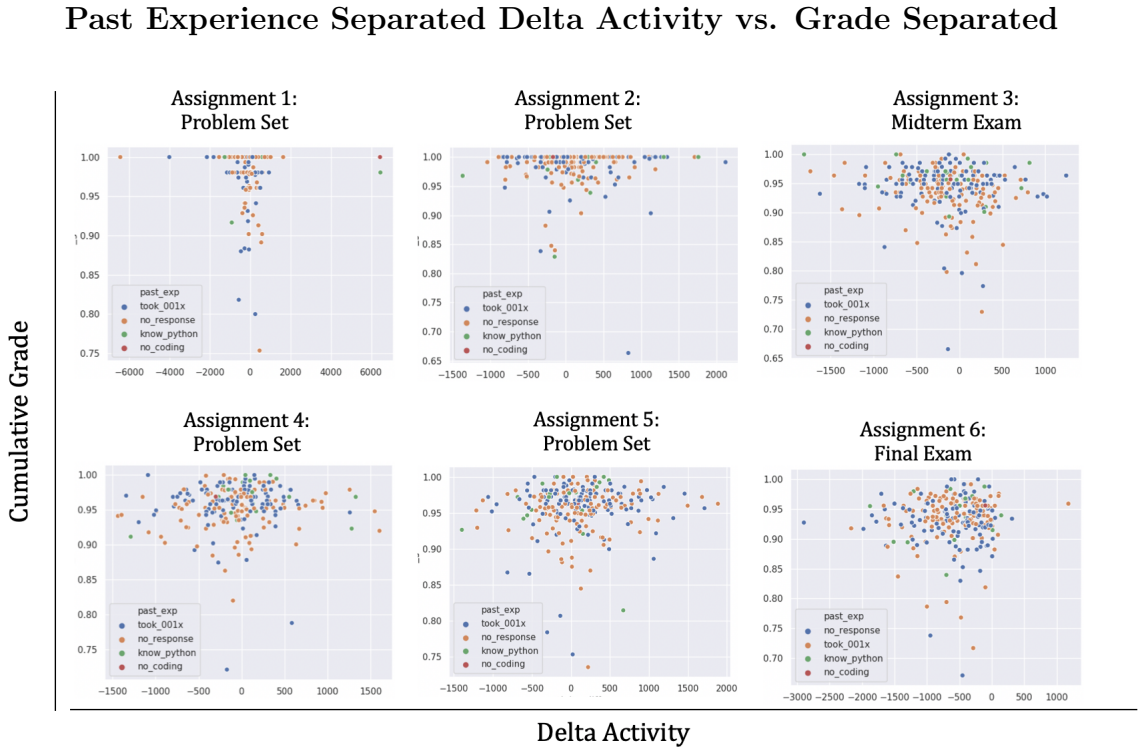

Figure 4-6: Spring 2017 6.00.1x Delta Activity vs. Grade Separated Into Past Experience

a good generalization of the population of student past experiences. In our case, as presented in Table 1.1.1, certified students make up less than 3% of all students. Therefore, students who have no prior knowledge could struggle more than students with prior programming knowledge, however, these students drop out quickly enough that we are unable to see their behavior. In chapter 6, we will look more into students who eventually drop out of the MOOCs and their behavior after receiving specific grades.

4.5 Summary of Key Observations

There are confounding variables when one observe behavior before and after an assessment. First, if the course material changes, i.e. flips in difficulty or underlying nature of knowledge (e.g. theoretical to procedural to analytical) or topic (mathematically oriented to language oriented). Setting these aside for future consideration,

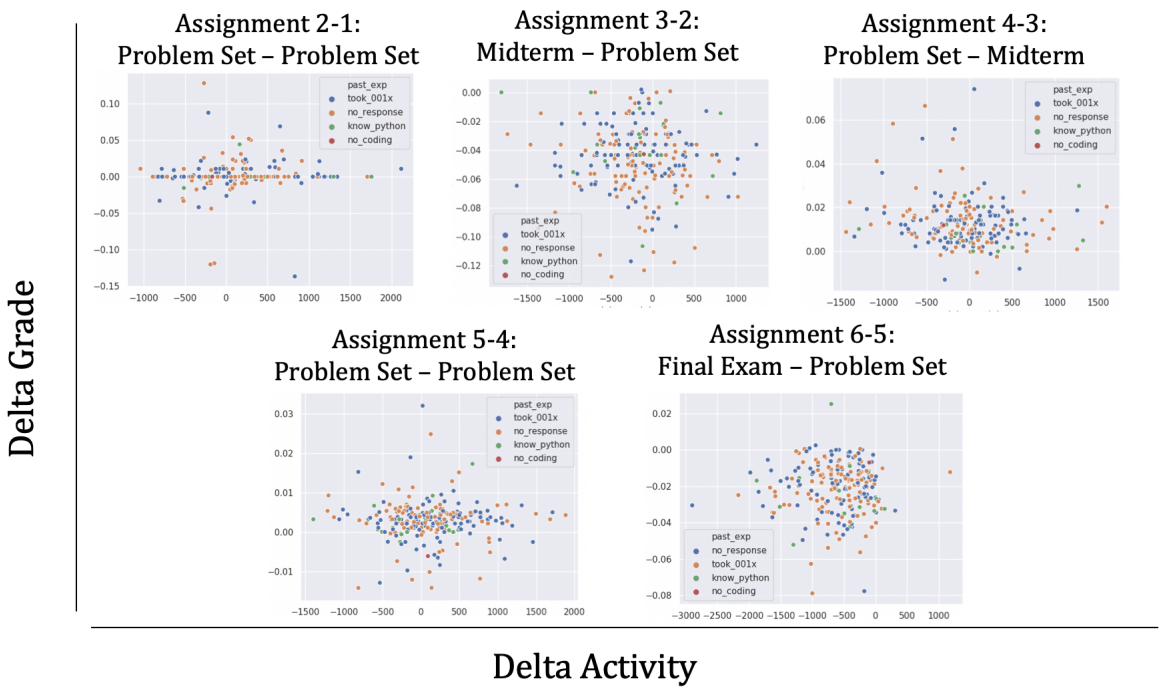

Figure 4-7: Spring 2017 6.00.2x Delta Activity vs. Delta Grade Separated Into Past Experience

all in all, in our study on the EdX MOOCs 6.00.1x and 6.00.2x, we were not able to find evidence to support that people change their behavior in accordance with their grades in the ways, accordance with our initial hypothesis. Instead we find more observations to the contrary and open the field to more investigations on lack of correlation between grades and learning behavior.

Chapter 5

Grade Trajectory Prediction with Machine Learning

In the previous section, we found no visible linear correlation between activity and grade, delta activity and grade, and delta activity and delta grade. In this section we will look into non-linear relationships between activity, past experience, and grade. Using logistic regression, tree-based classification, and neural networks, we train models to predict the grade trajectory of students.

5.1 Trajectory and Variable Selection

We create a binary variable called trajectory. Trajectory has a value of 1 if a student's cumulative grade stayed the same or became higher from its previous grade state. Trajectory has a value of 0 if a student's cumulative grade decreased from its previous grade state. We add the trajectory variable to all data sets with *cumulative grade* as a variable of interest (we do not add this to delta grade as it would be redundant). Trajectory will be the variable that we are trying to predict for the remainder of this analysis

Now that we have defined trajectory, we must also select which variables to use in our prediction. Because we are working with assignment-level granularity of data, we do not have very much detailed information. Therefore we develop basic models

that use all of our available data: before activity, after activity, assignment number, and previous experience. For the categorical previous experience, we create a binary variable for each category of past experience and have the variable equal 1 if the student had that prior experience and 0 otherwise.

5.2 Model Selection

In this section, we will discuss three different machine learning models that we used in predicting trajectory: logistic regression, decision trees, and neural networks.

Logistic regression is a powerful machine learning model used for supervised classification tasks [6]. Logistic regression is similar to linear regression in that we initialize a function and "fit" or linearly adjust the function based on the "training" of various data points. The main difference is that a non-linear function used to calculate the output and adjustment and each training iteration. We use the *LogisticRegression* class of python's sklearn module [1] to train and test our logistic regression implementation.

Decision trees are another model used for supervised classification tasks. Decision trees use predictor variables to construct a decision making sequence. For example given variables a, b, c, y where a, b, c are predictor variables and $y = \{m, n\}$ is the variable being predicted we can construct the following sequence: if $a < 1 and b > 2 and c < 4, y = m$. These sequences and their thresholds are then trained on existing predicted and predictor values for the best accuracy. For our implementation we use the *DecisionTreeClassifier* class of sklearn [2].

Neural networks are a versatile machine learning algorithm that can be used in a variety of tasks, including classification. Neural networks are similar to logistic and linear regression in that we select a model that can be "trained" on existing data points. Neural networks differ from regression in that there are multiple layers of intermediate nodes (which we can adjust) whose values we calculate before reaching

[1] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
[2] https://scikit-learn.org/stable/modules/tree.html

Best Trajectory Prediction Models' Accuracies

Model Type	Best Model Accuracy of Predicting Higher Grades	Best Model Accuracy of Predicting Lower Grades	Best Model Accuracy in Predicting Change in Grades
Logistic Regression Model	0.6361	0.5826	0.6124
Classification Tree-Based Model	0.7630	0.6714	0.7227
Neural Network Model	0.6500	0.6973	0.7896

Table 5.1: This table shows the cross-validated accuracies of the best trained model of each type.

the final output (or output layer). All calculations have a non-linear component, also differing from the regressions. We implement our neural network using the *MLPClassifier* class of sklearn [3].

5.3 Hyper-parameter Tuning and Best Results

Hyper-parameters are the values that specify the structure of the machine learning algorithms. These parameters allow us to exclude predictor variables and adjust the width and depth of the neural networks.

For all model types, we take the power set of our predictor variables and create models with each of those power sets. Additionally, for neural networks, we adjust the width and depth of the network to be some value $m \in [5, 50]$ and $n \in [5, 50]$.

Table 5.3 shows the test accuracy of the best of each type of model. Test accuracy refers to how well the model did on predicting the predicted variable from a test data set, different from but of the same population as the training data set.

Across all tests, the overwhelmingly most significant variable was "assignment number" suggesting that student grades follow a pattern specific to that class.

[3] https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html

5.4 Discussion and Conclusions

Through this analysis we again did not find an overwhelming correlation between grade trajectory and student behavior. The significance of "assignment number" could be due to the difficulty of certain assignments. For example, the midterm and final exams are more difficult than problem sets and therefore many students have lower grades after an exam. Similarly, problem sets after an exam will likely bring up grades because grades were lowered after an exam.

All in all, the analysis on certified students does not provide evidence of grade-correlated behavior change. In the next chapter, we will expand our data set to include high-activity not certified students in the form of *continuously participating* students.

Chapter 6

Continuously Participating vs. Certified Student Analysis

In the previous chapters, we found that we were not able to observe grade-correlated changes in behavior in certified students. We now broaden our field of analysis to *continuously participating* students (students who participate continuously up to specific assignments). We reapply our analysis from chapter 4 and compare the results with those from certified students and connected our observations to the existing self-regulated learning models from chapter 2.

6.1 Data Observations and Analysis

In this section, we discuss our observations and analysis on three comparisons between certified and continuously participating student activity and grade. Specifically we will look at the relationships between before/after activity vs. grade, delta activity vs. grade, and delta activity vs. delta grade.

6.1.1 Activity vs. Grade Analysis

Activity vs. grade shows the raw relationship between activity and grades. Before activity vs. grade shows the correlation between student activity and the grade they

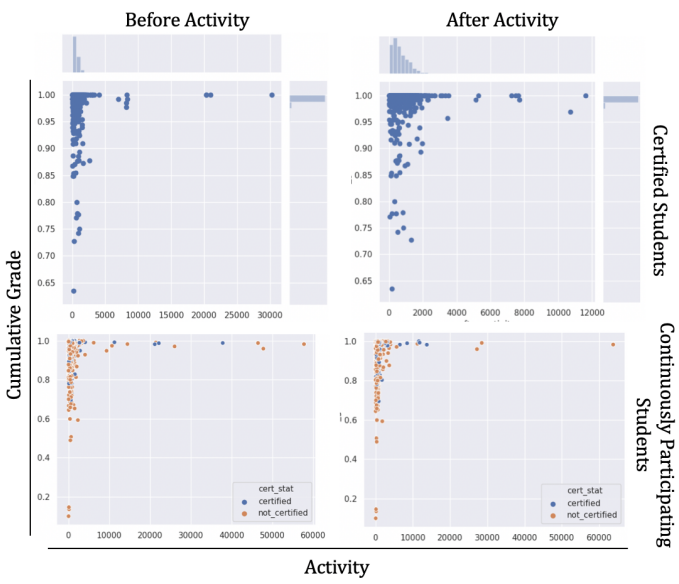

Figure 6-1: Depiction of the before and after activity vs. grade on certified and continuously participating students on problem set 2 of the Spring 2016 iteration of 6.00.1x.

receive with that level of activity, while after activity vs. grade shows the correlation between adjustments to activity and grade after a grade is given.

Figure 6-1 shows an example of the relationship between before (left column) and after (right column) activity and grade. The x and y axis are activity and cumulative grade respectively. The first row of graphs show the correlation for certified students while the second row depicts continuously participating (up to the second problem set of the Spring 2016 iteration of 6.00.1x) students. The two colorings of points in continuously participating graphs mark which students eventually receive certification and which do not: certified continuously participating students are blue while not certified students are orange. For more clarity in showing the distribution, for this and future graphs, we show students up to 3 standard deviations from the mean (excluding single outstanding outliers).

For both certified and continuously participating students, students with the highest levels of activity also have the highest grades. This is true for both before and activity as shown in figure 6-1. Figure 6-1 also shows that the before and after activity distributions are also very similar, this holds for all iterations of all courses.

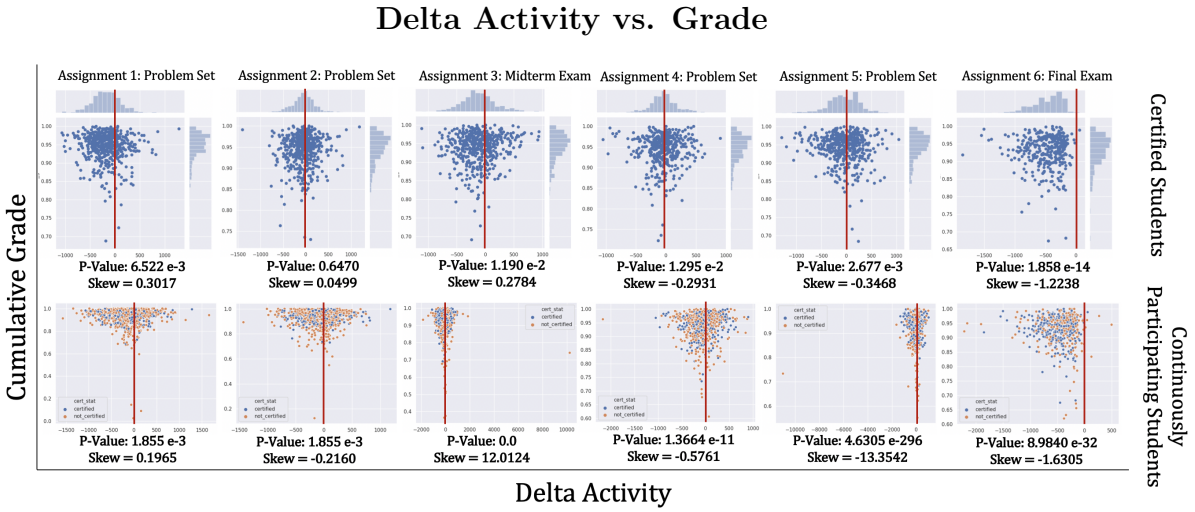

Figure 6-2: Depiction of delta activity vs. grade on certified and continuously participating students on the Spring 2016 iteration of 6.00.2x.

Next, we observe the grades of continuously participating students reach lower bounds than those of the certified students. Throughout our analysis, we find that continuously participating students who have a cumulative grade of below 0.5 do not eventually get their certification.

Lastly, we observe that nearly all certified and continuously participating students at the end of the course meet the passing threshold of the course (65 percent) without dropping their lowest assignment. In fact, 99 percent of continuously participating, certified students have a cumulative grade of above 70 percent in both 6.00.1x and 6.00.2x and more than 90 percent of continuously participating, certified student in all offering of 6.00.1x have a cumulative of above 90 percent.

6.1.2 Delta Activity vs. Grade

Delta activity vs. grade shows the difference in activity (before activity minus after activity) in relation to the grade received by students. This depicts to what extent students change their behavior as their grade changes.

Figure 6-2 shows the relationship and trends of delta activity vs. grade graphs for the Spring 2016 offering of 6.00.2x. The first row shows the transitions of delta activity vs. grade distributions for certified students in the Spring 2016 offering of 6.00.2x.

Delta Activity KS-Test P-values of Certified vs. Not Certified Continuously Participating

	Assignment Number							
Course	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
6.00.1x f2016	6.62 e-4	1.83 e-37	6.74 e-4	3.63 e-4	3.92 e-5	0.169	1.17 e-5	
6.00.2x f2016	0.538	0.015	3.47 e-4	0.289	0.012	4.55 e-4		
6.00.2x s2016	0.038	0.041	0.125	0.048	0.131	0.170		
6.00.1x s2017	0.044	0.054	0.098	0.049	0.095	0.034	0.049	0.067
6.00.2x s2017	0.089	0.131	0.086	0.051	0.050	0.086		

Table 6.1: P-values of the kolmogorov-smirnov test (KS-Test) for testing that the delta activity distribution of certified continuously participating students is the same as the distribution of not certified continuously participating students.

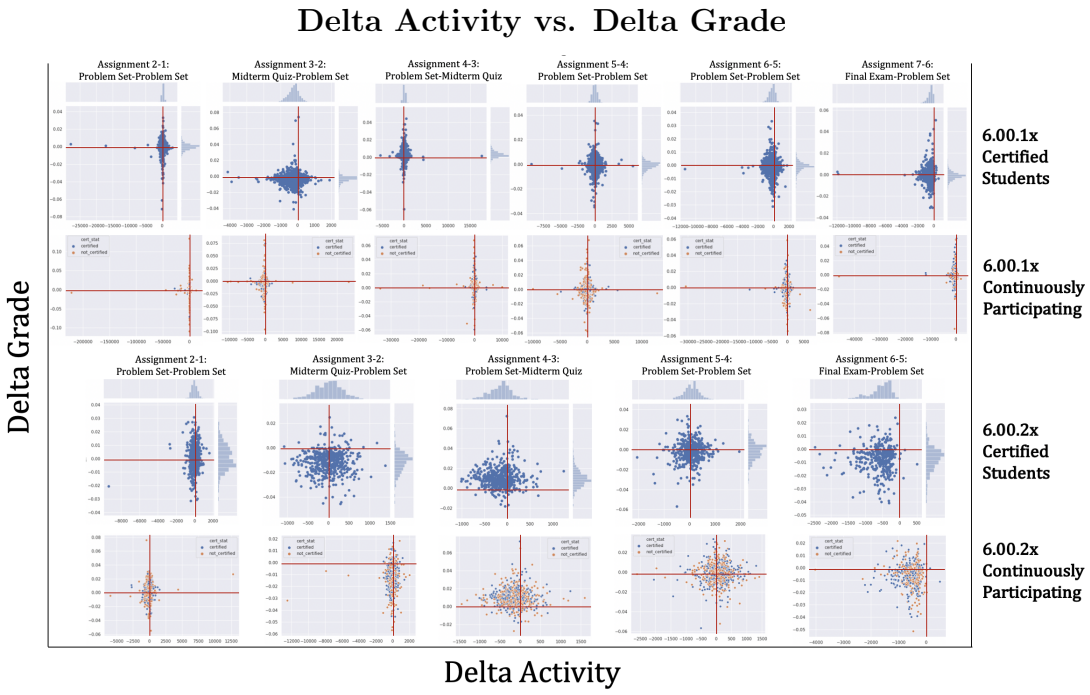

Figure 6-3: Depiction of delta activity vs. delta grade on certified and continuously participating students on the Fall 2016 iterations of 6.00.1x and 6.00.2x.

Delta Grade KS-Test P-values of Certified vs. Not Certified Continuously Participating

	Assignment Number						
Course	**1**	**2**	**3**	**4**	**5**	**6**	**7**
6.00.1x f2016	0.116	0.313	0.549	0.160	0.822	0.313	
6.00.2x f2016	0.479	0.347	0.715	0.804	0.822		
6.00.2x s2016	0.050	0.815	0.811	0.064	0.779		
6.00.1x s2017	0.015	0.637	0.789	0.840	0.454	0.684	0.549
6.00.2x s2017	0.121	4.29 e-3	0.169	0.814	0.850	0.354	

Table 6.2: P-values of the kolmogorov-smirnov test (KS-Test) for testing that the delta grade distribution of certified continuously participating students is the same as the distribution of not certified continuously participating students.

The second row shows this relationship with continuously participating students. The bottom (x-axis) of all graphs is delta activity and the left (y-axis) of all graphs is cumulative grade. Additionally, we add the skew and p-values found for a normality test of the delta activity distribution. These normality tests have a null hypothesis that the delta activity distribution is normal. We use the `scipy.stats.normaltest` function [1] to aid in our analysis.

The red lines on each graph marks the line x=0, where the before and after activity of a student are the same. Points on the left of the red line indicate students who do less activity after receiving a certain grade, while points to the right of the red line are students who do more activity after getting a certain grade. For most assignments (except for the last assignment at the end of the course and other outliers), the activity difference is centered close to x=0 and fairly symmetric. We observe this trend in all offerings of both courses and in both certified and continuously participating students as seen in Figure 6-2. This observation suggests that, as a whole, the distribution of students' activity change does not linearly correlate with grade received.

From the normality tests, we can say that with high confidence (greater than 95 percent), for *most* delta activity distributions, that we reject the null hypothesis that the delta activity distributions are normal. In the skew tests, we see that the skew of the delta activity distributions of certified students is much smaller in magnitude than

[1]://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.normaltest.html

those of the continuously participating students. This is partly due to outlying values of students who drop out or do not receive their certification. However, some outlying values are due to drastic changes in behavior. In row 2 column 3: Midterm Exam, we have a high positive skew due to increases in activity in *not certified* continuously participating students. These students had cumulative grades that range between just below 70 percent to above 90 percent, with most scores being on the lower side. This behavior could be explained by "students increasing their activity due to receiving a grade below their expectations." However, it is notable that the most dramatic activity increases at each grade level were *not certified* students, suggesting that these changes (and their results) did not encourage the students to stay in the course but rather eventually drop it. Interestingly, higher positive skews are not found in the delta activity vs. grade graphs of certified students, suggesting that certified students who receive poor grades do not try to do more activity and change their learning behavior to increase their grades.

After this, we analyze the difference in the distributions of certified continuously participating and not certified continuously participating students. Using the Kolmogorov-Smirnov test from the scipy python package [2], we find the p-values of the null hypothesis that the distributions of the two samples are the same. Table 6.1.1 shows the p-values for all assignments of all offerings except 6.00.1x s2016. 6.00.1x of Spring 2016 did not have enough values to be analyzed using the python package. We observe that although some offerings have all assignments test with small p-values (such as 6.00.1x f2016), others have high p-values (such as 6.00.2x s2017). For 6.00.2x s2017, at a 5% alpha level, we would not be able to reject the hypothesis that certified continuously participating and not certified continuously participating students are from the same distribution. Due to this discrepancy, we cannot draw any meaningful conclusions from this analysis.

Finally, as grades decrease, we see less variation in the magnitude of delta activity. This can be interpreted as students with higher grades having greater changes in behavior than students with lower grades.

[2] https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.ks_2samp.html

6.1.3 Delta Activity vs. Delta Grade

Lastly, we look at delta activity vs. delta grade, which shows the correlation between changes in activity and changes in grade. To reiterate, delta activity is the before activity minus the after activity when a grade is received and delta grade is the change in cumulative grade when a grade is received.

Figure 6-3 shows the delta activity vs. delta grade graphs of the Fall 2016 offerings of 6.00.1x and 6.00.2x. The bottom x-axis of all graphs shows delta activity and the left y-axis shows delta grade. The right row labels show the row's course and student set. The labels above the first and third rows show the assignment numbers' grades that delta grade was taken from and the type of assignment (test or problem set).

The red lines on each graph are the lines x=0, y=0. x=0 separates students who increased their activity (right half) from students who decreased their activity (left half). y=0 separates students whose grade increased (top half) and those whose grade decreased (bottom half). Therefore we make the following interpretations: the top-left quadrant are students who decreased their activity after their grade increased, the top-right quadrant ware students who increased their activity after their grade increased, the bottom-left quadrant are students who decreased their activity after their grade decreased, and the bottom-right quadrant are students who increased their activity as their grade decreased. For all graphs, we do not see strong evidence that certified or continuously participating students change their activity when specific grades are received nor do we see students with large changes in grade be attributed to specific changes in activity. This trend holds for all offering of both courses.

For 6.00.1x, for both certified and continuously participating students, we see a trend of delta activity vs. grade graphs centering along x=0, y=0. This suggests that the largest variations in grade occur with little or no change in activity, while the largest changes in activity correlate with little or no change in grade. We call this formation the "star" shape which is prevalent through all offerings of 6.00.1x and shown in the first 2 rows of Figure 6-3.

For 6.00.2x, although there are still occurrences of the "star" shape, they are less

prevalent and well defined. For the example, the second and third columns of the 6.00.2x rows are centered below and above y=0. The trend of less "star"-like shapes persists through all offerings of 6.00.2x.

Next, we observe that there is higher magnitude, in all directions, of delta grade and delta activity of continuously participating students compared to certified students. In rows 2 and 4 from the top, the orange markers indicating *not certified* continuously participating students, can be seen on the outer corners of the distributions in all directions.

Lastly, we analyze the delta grade distribution of our delta activity vs. delta grade graphs. Table 6.1.2 shows the p-values of a Kolmogorov-Smirnov test on the delta grade distributions of certified continuously participating students and not certified continuously participating students. The null hypothesis was that the two samples came from the same distribution, in other words that the delta grade distributions of certified continuously participating students and not certified continuously participating students are the same. We use the python scipy function ks_2samp [3]. Additionally, there were not enough data to test Spring 2016 6.00.1x with this function, therefore, it is not included in this analysis. Table 6.1.2 shows that for most assignments of all offerings, we cannot reject the null hypothesis that the distributions are the same, as most p-values are more than 0.05.

6.2 Discussion

In our analysis of activity vs. grade, we find that students with the highest levels of activity also have high grades with lower students with lower grades drop out. This is reasonable as we expect to see students to be able to succeed with hard work. Low activity individuals have a wide spread of grades, suggesting that other factors such as past experience and inherent comprehension contribute can take precedence at similar, lower activity levels. We also find that continuously participating students with cumulative grades lower than 0.5 eventually drop out

[3] https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.ks$_2$*samp.html*

or do not receive certification. This is consistent with the results by Halawa et al. [8]. In our case however, this is also a reasonable result because one of the criteria for certification is a final grade of 0.65. If a student mid-course has a grade already severely below this threshold, it is reasonable for the student to conclude that they do not sufficiently understand the course material to pass or receive certification.

In our delta activity vs. grade analysis, we see that students generally do not drastically change their activity after getting a specific grade. Furthermore, the largest activity changes come from students who do not become certified. This could be because of the expectations and desired goals of the student. Students with the highest goals will strive harder and more drastically change their learning behavior to meet those goals. At the same time, inability to meet exceedingly high goals despite the student's increased efforts can be a driver of dropout [1]. On the other hand, students with consistently lower but passing grades do not drastically change their activity. This could be because these students are not able to dedicate as much time to the course and can only keep up their steady progress to become certified or learn their desired material. It could also be because the students have lower goals that they are consistently able to meet.

The delta grade vs. delta activity analysis shows differences in trends between courses. The dispersion of the star shape in 6.00.2x is because grades were able to be lowered and raised in the student distribution as a whole (through a difficult exam and subsequent easier problem set). This can be attributed to the hardness of the assignment and the difficulty of the course as a whole. As 6.00.2x teaches more advanced topics than 6.00.1x, it would be reasonable that the exams and homework assignments are more challenging.

Additionally, we found that we cannot reject that certified and not certified participating students are from the same distribution of delta grade. This suggests that changes in grade are very similar for the two groups. Future work can help us understand why this occurs.

Lastly, delta activity vs. delta grade graphs showed higher delta grade and delta activity magnitudes for continuously participating students. This means that con-

tinuously participating students are have greater mobility in grade and activity, and therefore learning behavior. Large changes in activity are likely correlated dropout and reasons discussed previously. Higher grade increases could be because continuously participating students had lower grades to begin with, and therefore had more they could improve. Higher grade decreases could be due to selection bias: because these students' grades decreased by a great amount, the students eventually dropped out or could not receive certification.

6.3 Conclusion

Our initial hypothesis that students, and most prominently certified students, in MOOCs adjust their activity to pass the course does not seem to be supported by our observations. Although there were outliers in delta activity vs. grade continuously participating that do seem to support our hypothesis, it cannot be seen in the student distributions as a whole; furthermore, the outliers eventually dropped the course. One possible explanation of this could be the misconceptions of students in how or what they change in their learning behavior and how much they can achieve with that change. This issue was raised by Butler and Winne in their study about misinformation in learning and how it affects feedback [1]. Another explanation could be the grade feedback does not contribute to significantly to learning behaviour and that other intrinsic properties of the student take precedence.

Again, although we did not find strong linear correlations between grade and activity, we did observe the correlation between low grades and dropout.

Chapter 7

Principle Conclusions and Future Work

In both our certified students and continuously participating students analysis of grade correlated changes in behavior, we found little to no correlation between grade and student behavior. Furthermore, we also found poor results in grade prediction using activity. This could suggest an inherent disconnect between grades and how students behave and learn in MOOCs. This would in turn suggest that the current grading system and certification system, which are modeled from traditional school grading, are not suited for online courses. This opens the discussion of whether it is then valuable at all for grades to be the main certification criteria. Other future work in this field could include analyzing all students with different past experiences and looking into how grades and other feedback mechanisms effect these groups of students. It would also be interesting to look into delta grade and understand why changes in grade for not certified and certified continuously participating students are not different.

We hope that this work will inspire future work in this field, and in turn improve MOOCs for generations of current and future students.

Bibliography

[1] Deborah L. Butler and Philip H. Winne. Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3):245–281, 1995.

[2] Charles Carver and Michael Scheier. Origins and functions of positive and negative affect: A control-process view. *Psychological Review*, 97:19–35, 01 1990.

[3] T. Daradoumis, R. Bassi, F. Xhafa, and S. Caball. A review on massive e-learning (mooc) design, delivery and assessment. In *2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, pages 208–213, Oct 2013.

[4] Paula de Barba, Gregor Kennedy, and Mary Ainley. The role of students' motivation and participation in predicting performance on a mooc. *Journal of Computer Assisted Learning*, 32:218–231, 03 2016.

[5] Mucong Ding, Yanbang Wang, Erik Hemberg, and Una-May O'Reilly. Transfer learning using representation learning in massive online open courses. *arXiv preprint arXiv:1812.05043*, 2018.

[6] Stephan Dreiseitl and Lucila Ohno-Machado. Logistic regression and artificial neural network classification models: A methodology review. *J. of Biomedical Informatics*, 35(5/6):352–359, October 2002.

[7] 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.

[8] S Halawa, D Greene, and J Mitchell. Dropout prediction in moocs using learner activity features. *Proceedings of the Second European MOOC Stakeholder Summit*, 37:7–16, 01 2014.

[9] Daniel T. Hickey. Massive open online courses. In Kylie peppler, editor, *The SAGE Encyclopedia of Out-ofSchool Learning*, pages 460–462, Thousand Oaks, 2017. SAGE Publications, Inc.

[10] Suhang Jiang, Mark Warschauer, Adrienne E. Williams, and Katerina Schenke. Predicting mooc performance with week 1 behavior.

[11] Katy Jordan. Massive open online course completion rates revisited: Assessment, length and attrition. *The International Review of Research in Open and Distributed Learning*, 16(3), 2015.

[12] 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.

[13] Marius Kloft, Felix Stiehler, Zhilin Zheng, and Niels Pinkwart. Predicting mooc dropout over weeks using machine learning methods. pages 60–65, 01 2014.

[14] Richard Magill. The influence of augmented feedback on skill learning depends on characteristics of the skill and the learner. *Quest*, 46:314–327, 08 1994.

[15] Joyce B. Main and Ben Ost. The impact of letter grades on student effort, course selection, and major choice: A regression-discontinuity analysis. *The Journal of Economic Education*, 45(1):1–10, 2014.

[16] Sinclair Jane Onah, Daniel F. O. and Russell Boyatt. Dropout rates of massive open online courses : behavioural patterns. In *EDULEARN*, volume 14, pages 5825–5834, Barcelona, Spain, jul. The Warwick Research Archive Portal, The University of Warwick.

[17] Zhiyun Ren, Huzefa Rangwala, and Aditya Johri. Predicting performance on MOOC assessments using multi-regression models. *CoRR*, abs/1605.02269, 2016.

[18] Jacob Whitehill, Kiran Mohan, Daniel T. Seaton, Yigal Rosen, and Dustin Tingley. Delving deeper into mooc student dropout prediction. *CoRR*, abs/1702.06404, 2017.

[19] Jacob Whitehill, Joseph Williams, Glenn Lopez, Cody Austun Coleman, and Justin Reich. Beyond prediction: First steps toward automatic intervention in mooc student stopout. *SSRN Electronic Journal*, 01 2015.

[20] Diyi Yang, Tanmay Sinha, David Adamson, and Carolyn Penstein Rosé. turn on , tune in , drop out : Anticipating student dropouts in massive open online courses. 2013.

[21] Cheng Ye and Gautam Biswas. Early prediction of student dropout and performance in moocs using higher granularity temporal information. *Journal of Learning Analytics*, 1:169–172, 01 2014.