Student Interaction with Online Pre-Lecture Videos and how it Influences Grades

by

Allyson Petersen

May 12, 2017

Advisor: Dr. KC Walsh

Department of Physics

Oregon State University

Abstract

This study examines how students were interacting with pre-lecture videos via the BoxSand website. Students of the Fall 2015 and Fall 2016 introductory physics courses at Oregon State University were studied. The course was taught as a flipped classroom in which the instructional material was provided via BoxSand. BoxSand is a website designed to provide course material for the physics course studied, and it records how students interact with the site. The purposes of the study were to show how BoxSand can be used to improve learning among introductory physics students and to explore how students interacted with pre-lecture videos.

Educational data mining was performed on data collected by BoxSand, with analysis focusing on the data generated by students clicking on links to pre-lecture videos. Formulas were written in Excel to analyze the datasets.

The major results from the Fall 2015 dataset were that students tended to increase the number of videos they watched before each exam and they typically watched videos for the actual length of the video. This second result is important because the website tracks how long a student stays on a page, it does not track how the students interact with the video once they are on that page.

The most important results for the Fall 2016 dataset were that watching more pre-lecture videos tended to correlate with better grades and the students who increased the percent of videos they watched had the best chance of improving their grades in the next exam.

The methods used in this project can be expanded to analyze how students use other parts of the website to study or learn course material. When more terms of data become available for analysis, a larger picture of how students change their study habits throughout the introductory physics sequence can be made.

Acknowledgements

I am grateful for all the help I've had throughout this project. First, thank you to my advisor, Dr. KC Walsh, for his continued support and guidance. I also thank the rest of the BoxSand research team, Ryan Scheirer and Tallon Korn, for all their feedback and advice with my research. Finally, thanks to the students of my thesis class for helping edit and providing feedback on my thesis.

Table of Contents

List of Tables and Figures

1. Introduction

1.1 Motivation and Objective

Physics professors expect students to be highly engaged with course material, perhaps more than students are used to if they are new to studying physics. Professors stress the importance of students being more involved, but students often keep their old study habits, which can result in grades that are lower than a student expects. Physics students are taught to look at evidence, rather than simply accepting something as factual. Perhaps if students are provided some evidence on what behaviors correlated with gains in grades, they would be more likely to adjust their existing study habits.

This project aims to provide students with evidence of how the amount a student is engaged with the course material can affect grades. We hope that students will be more willing to adjust their study habits when presented with this information.

1.2 Background

This project uses the website BoxSand to correlate student engagement with course material with students' grades. The method used to analyze the large sets of data is relationship mining. Learning Analytics is used to relate the results to the students in the class.

BoxSand

BoxSand is an open-source website used for teaching introductory physics courses. It is designed to be a sufficient replacement for textbooks, and therefore save students money. The website tracks the actions of students while they interact with the website. The data collected by the website provides instructors insight into the study habits of their students. BoxSand can also facilitate a flipped classroom by giving the students easy access to course material. A flipped classroom is taught opposite from a typical course; students learn most of the course material outside of class so that class time is freed for working examples, answering questions, and for other material review.

The BoxSand website was designed and created by Dr. KC Walsh with help from undergraduate and graduate students. Dr. Walsh currently teaches introductory physics courses at OSU using the BoxSand website with the flipped classroom design. The website was introduced to the students at the beginning of the Fall 2016 course. Permission to keep track of data was granted by students who were willing to take part in the project. The site records data that is generated by students engaging in the site. If a student performs an activity such as clicking on a link, watching an embedded video, or clicking on a menu drop-down button, the site keeps track of it. It records data for each student and keeps track of the time a page was loaded and when it was left, as well as where the student was directed there from.

At the end of the term, the data was accessed by Dr. Henri Jansen, a physics professor and head advisor of the physics department. He was not involved with the courses being studied and is the only person with permission to view the raw data. Dr. Jansen removed the identifying information in the data and replaced them with unidentifiable markers like "USER 124"; this was done to protect the privacy of the students. In January, Dr. Jansen provided Dr. Walsh with this edited information, who in turn provided the rest of the research team with the necessary data for data analysis. The dataset listed usernames, date and time stamps, user grades, pages visited, and more.

Educational Data Mining

Educational data mining (EDM) involves analyzing large amounts of educational data with the goal of improving education [1]. EDM deals with dataseets that are too large to analyze with other methods. Some common methods that EDM uses include prediction models, structure discovery, and relationship mining [1]. The prediction model method aims to make predictions for a variable in the data based on other variables in the data [1]. Structure discovery aims to find structure in the data without previously knowing what should be found [1] The main model that is used in this project is relationship mining which aims to determine relationships among the many variables in the data [1]. Future work will aim to make prediction models using the data.

In the literature, educational data mining has been used to analyze both pre-existing and new data [2]. Some of the research that used pre-existing data worked with data that is typically collected by a university, like grades, gender, area of study, and high school grades. Researchers using educational data mining use their findings to predict which students are at risk of dropping out and look for a solution to decrease this risk. These studies tend to have the objective of testing the effectiveness of EDM before going to the effort of taking new and large data sets.

The research that is more relevant to this project are the studies that take new data with the intention of using EDM methods to analyze it. First, large amounts of data relevant to the study are taken. Then EDM methods are used to analyze the data.

Learning Analytics

Learning Analytics (LA) is another method for analyzing data. Learning analytics concerns the measurement, collection, analysis, and reporting of data [3]. Data is extracted from an educational environment and then processed. A common method used by LA is content analysis. In this project, Learning Analytics is used in relating the outcomes we get to the students in the class. Sometimes it is necessary to look at the data while considering the individual student. The data may show a trend that doesn't agree with the predicted one. We understand that individuals have different study and learning habits which may not fit with what we assume or observe in the data for that person.

BoxSand Data Analysis

The BoxSand website records how students interact with the site's content, like clicking on links and watching videos, when the students are logged on. Large sets of data are generated from these interactions. Methods of Educational Data Mining and Learning Analytics are used to analyze these data sets.

First, the important data is selected from the larger set of data; the largest set of data includes all of the information recorded by the website except for student names and IDs. Then, the EDM relationship mining method is used to find relationships and correlations in the data. LA methods are used to find trends in the data, draw conclusions from these trends, and then make improvements based on the results. EDM and LA methods are an important part of the project; the information gained from them will suggest how the BoxSand website can be improved to facilitate student learning.

Another source of data for the project was a survey given to the students each term. The survey questioned students on the BoxSand website, particularly about its usefulness and ease of use. Survey responses show how the site can be improved, but also provides more data to study. The survey gives insight into what the students are thinking; important information that cannot be collected by the website itself.

2. Fall 2015 BoxSand Dataset Analysis

2.1 Introduction

Data was analyzed from the Fall 2015 class. The purpose was to get an overall view of how the class interacts with the course videos and to start forming ideas on how to analyze the data based on what is possible from the data collected.

Only three months of data collection were available; these were October, November, and December of 2015. The data was analyzed using Excel. Formulas were written to analyze the data, depending on what information needed extracted from the data set; the formulas written included countifs, averageifs, sumifs, and vlookup statements.

The information provided on the dataset included what webpage (on the BoxSand website) and when it was visited, where the student was directed from (or previous page loaded), and which student it was that visited the page. As explained previously, the student names and ID's have been removed and replaced with labels that can only distinguish students from one another in the dataset, but cannot be used to identify actual students.

2.2 Methods

The minutes spent on a given webpage was determined by first sorting the dataset by student. For each student, the data were still organized by time. Then, the timestamp from the next page was subtracted from the timestamp of the page in question; this was applied to the entire dataset. Any large difference indicated a change in day or a general break in activity with the site. The data analysis using this information had to accommodate this by not including the large times, or only choosing the amount of time of interest.

We can only assume that students watch the videos when visiting a video page. The dataset does not provide information on how the student interacted with the video (pause, play, fast forward, etc.) To get the time a student interacts with a video page, the timestamp of when a video page was visited is subtracted from the timestamp of the next page visited.

2.3 Results and Analysis

The average number of videos watched by the entire class (all sections) each month were determined. A formula was written that counted the number of video pages visited if the page was loaded during the desired month. The formula also included if statements to exclude page times of very short or very long times. To normalize the data, the monthly video totals were divided by the number of days the data was collected each month, since November had the largest amount; the term finished early December, and the data collection didn't start October 1st. Figure 2.1 displays the results.

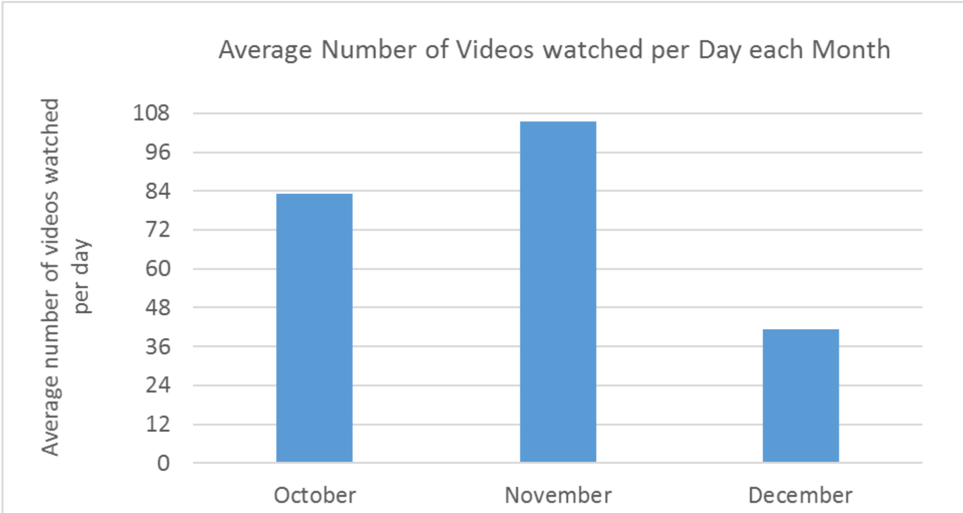

Figure 2.1: The average number of videos watched each day during each month was determined. The data was normalized by the number of days that data was collected in each month.

The average number of videos watched per day was much higher in November than the other months. December's lower average was expected, since there tends to be less content delivery, and more time spent studying because it is the end of the term and students are preparing for the exam. October was expected to have an average closer to November's average. This discrepancy is possibly due to not having data for the entire month; even though the data were normalized, a significant portion of the videos may have been watched before data collection started and this wouldn't be accounted for with the normalization. More months of complete data collection should be used for further analyses. The actual number of videos assigned to be watched should also be considered; there may just not be as many videos assigned in October, for instance.

Breaking this analysis down further, the number of times videos were loaded each day was analyzed. The method of analysis was similar; the number of videos loaded each day was counted, but the data didn't need to be normalized. The results are presented in Figure 2.2.

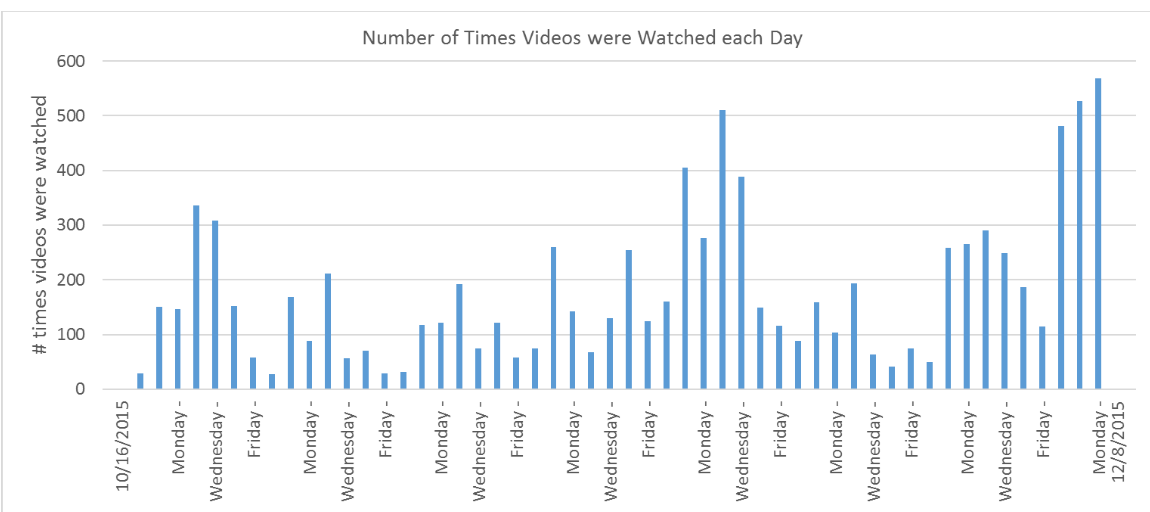

Figure 2.2: The total number of times videos were watched each day throughout the course. Each bar indicates a day, only Mondays, Wednesdays, and Fridays are labeled.

The results given in Figure 2.2 shows a general increase in number of videos watched at the beginning of the week (near Mondays). The videos were assigned as pre-lecture videos to be watched before class on Monday, Wednesday, or Friday, depending on the video. This led to the expectation of peaks on or before those days, with activity dropping off on the weekends (but starting to peak on Sunday).

Specific videos were analyzed to start looking at how watching videos and earned grades are correlated. Videos of various lengths were selected, including lengths of approximately 2, 4, 8, 10, and 44 minutes. Figures 2.3 and 2.4 show the results for a 1 minute and a 4-minute video. The chosen plots show the common trends found in the analysis. Students in the B group tended to watch the videos at a higher rate than other students. Also observed is that D students occasionally watched the videos more than the A students. Two possible reasons for this include: students with A grades are already familiar with the course content and don't need to study as much, and the number of D students is small, so variations in study habits have a large impact on the data analysis.

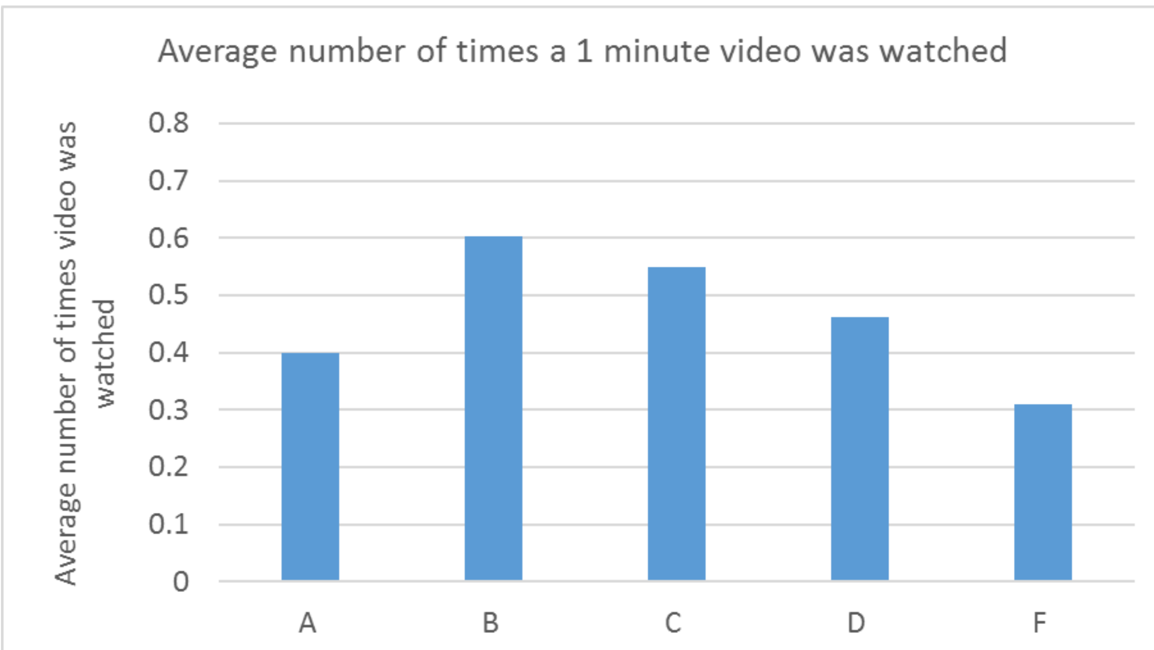

Figure 2.3: A video approximately 1 minute long was chosen and the number of times each student watched it was counted. Students were grouped by course grade and the average number of times the group watched the video was determined.

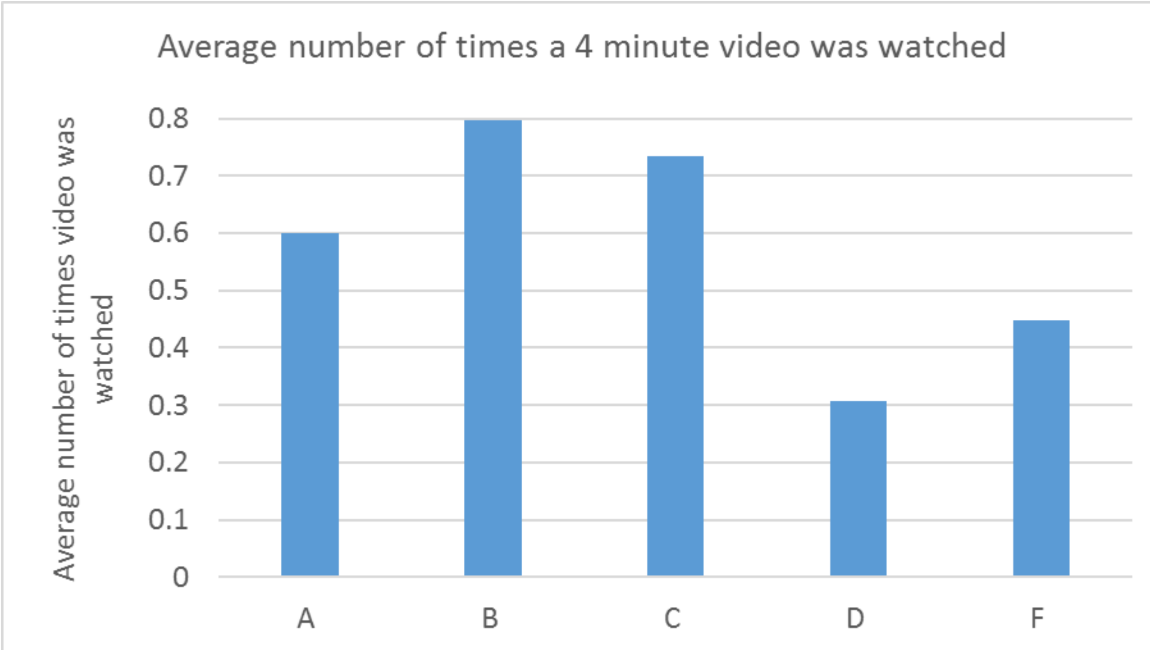

Figure 2.4: A video approximately 4 minutes long was chosen and the number of times each student watched it was counted. Students were grouped by course grade and the average number of times the group watched the video was determined.

Videos have the option to be paused, resumed, the speed can be changed (0.5x, 1x, 1.5x, and 2x the speed), and it can be fast-forwarded or rewound. To explore how students watched the content videos, the amount of time certain videos was watched was determined; the results are provided in Figures 2.5 and 2.6.

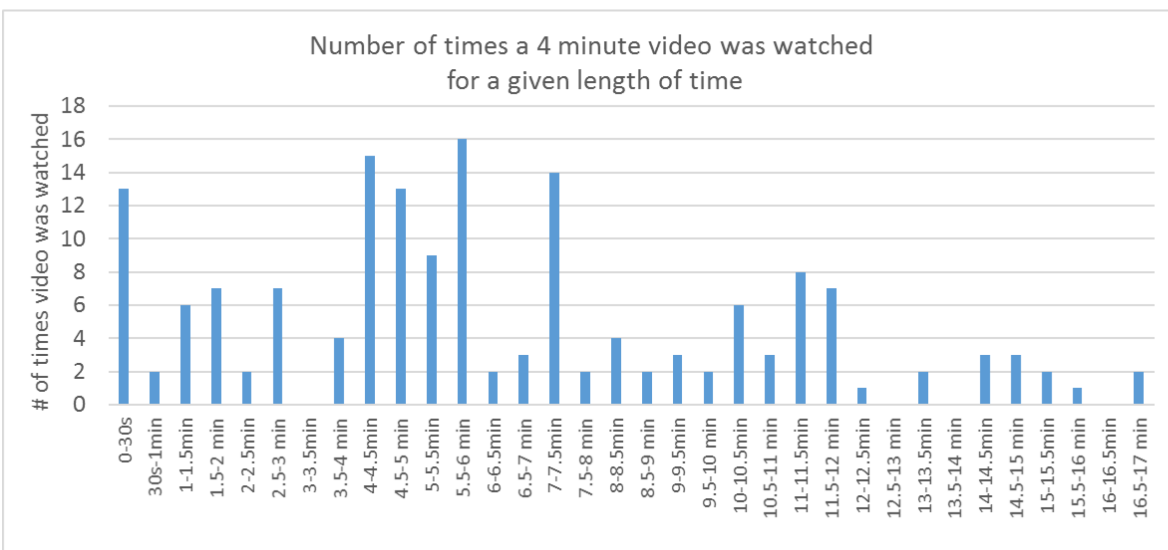

Figure 2.5: A 4-minute long video was selected and the number of times it was watched, by any or all students, was determined for certain lengths of time. The video was watched 164 times total and was about 3.87 minutes long.

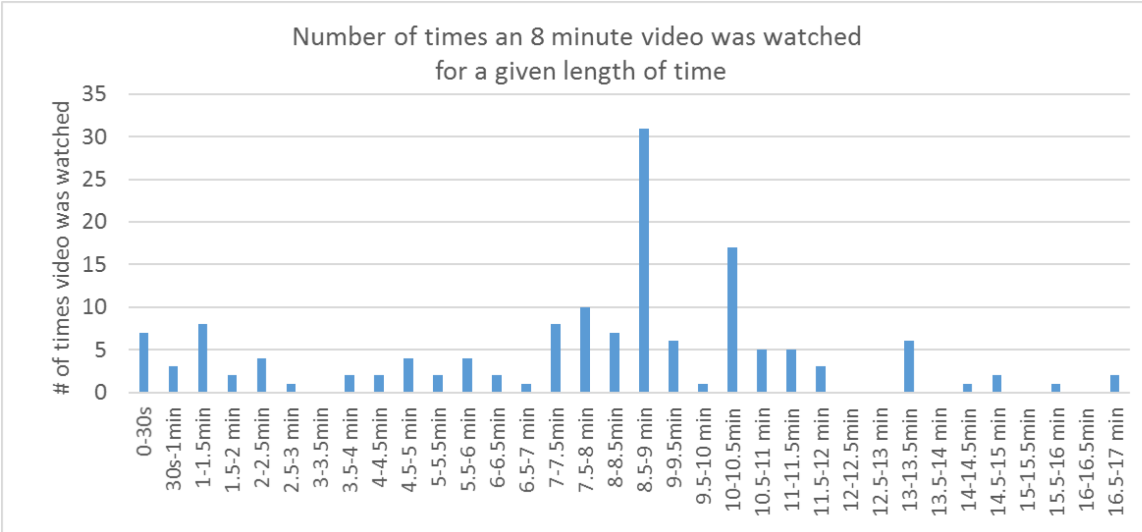

Figure 2.6: A 8-minute long video was selected and the number of times it was watched, by any or all students, was determined for certain lengths of time. The video was watched 147 times and was 8.22 minutes long.

The results displayed in Figure 2.5 and 2.6 suggest that students typically watch a video at normal speed for the duration of the video. This was the desired result; since the videos are intended to be pre-lecture content, students may not be as likely to interact with the videos (i.e. pause and rewind) as was initially expected. The highest number of viewings were slightly above the video's actual runtime; this suggests that students may interact with the videos, such as pausing to take notes or rewinding, but it could just be due to how the students are tracked by the website. The length of time a student spends watching a video is determined by when they open the page with the video, and when they visit another page on the BoxSand website. This method introduces error, even when a student quickly switches to a new page (this error is why later analysis of the data uses page visits rather than time spent on pages).

2.4 Summary

The Fall 2015 dataset was analyzed to get an overview of how students interacted with videos. Students seemed to be watching more videos in November, but with only three months of data, this could be saying more about the number of videos assigned in each month than about the students' decisions to watch videos. Breaking down by day the number of videos watched showed trends in how students prepared for lecture and exams. A large increase in the number of videos watched occurred near each exam. Increases in videos watched also occurred before the lecture days. How long a student actually watched a video was not available, we assumed that when students were done watching a video they immediately visited another page on the BoxSand website. The length of time students spent on specific videos was found to closely match the actual length of the video a majority of the time.

3. Fall 2016 BoxSand Dataset Analysis

3.1 Introduction/background/motivation

The Fall 2016 class was introduced to the BoxSand website at the beginning of the course and were expected to use it throughout the entire course. This dataset was de-identified and provided to the research group; it provided the same information as the dataset for the 2015 class. The purpose of analyzing this dataset was to look not only how is the class interacting with videos, but also as individuals, how a student interacts with videos. Specifically, which videos are watched, for how long, when, by who, and how do these impact grades and then change with subsequent exams.

The larger goal of analyzing this data was to follow the students through the following courses (Winter and Spring terms) for the physics sequence and see how their site interactions change over the large scale. The video data that this paper focuses on will be less reliable for future comparison due to an error in the data collection that was discovered after the term was finished. The website was not collecting data when students right clicked on a link to open it in a new tab if the link sent the student to a non-BoxSand site. Only about a third of the videos for Fall 2016 were located on-site, so some of the interaction data is missing for the other two-thirds of the term. It would not be easy or, perhaps, possible to just remove students who right-click to open new tabs because students aren’t consistent on which method they use (left vs right click). The best way to overcome this problem is to only analyze the data from the first third, and last third of the term where right-clicks were not able to be tracked. Note that the term was split into thirds (for analysis only), depending on where the pre-lecture videos were linked to, offsite versus onsite, not by days or content taught.

Much of the analysis performed was done as stepping stones to get to the desired analysis or plot. For example, scatterplots were typically created before grouping data into bins to create a bar plot that is easier to read and interpret; the scatter plot helps to check that the bar plot is accurately displaying the data.

The class grade (or course grade) includes points earned from going to lab, lecture, and from homework. Some of these points don’t represent a student’s understanding of the material; a student may have a lot of help on homework assignments, so even those points can’t accurately indicate a student’s understanding. The exam grades (from two midterms and the final) are more accurate indicators of learning. For analysis using exam grade as a variable, the exam grade is the weighted average of the two midterms and the final grades.

There were two types of pre-lecture videos: suggested and required. The suggested videos, as the name implies, were suggested to the students to be watched for the corresponding class day, but they were not necessary or required to be watched. The required videos were pre-lecture videos that students were expected to watch before the corresponding class.

The ph201 class was graded as follows. Certain students were not counted in certain analyses because the grade information for these students were not available. All of the videos used for analysis were created by Dr. KC Walsh; videos that linked to other sites, like YouTube, were not counted or analyzed. The term videos in this paper, unless noted otherwise, refers to pre-lecture videos that students were expected to watch before the lecture day corresponding to the video.

Anytime restrictions were put on counting videos watched, some videos were not recorded as having been watched; an example of these restrictions is not counting if the next timestamp was longer than two hours from the video page being loaded. The restrictions were tested to make sure the trends were not affected by them; different values were tested until an acceptable one was found that did not affect the trend, but made sense to have. Restrictions placed on the data did affect the numbers, including: the number of videos watched and the time spent on the site or certain activities; however, the analysis is most concerned with patterns and trends in the data, less on the numbers, so these restrictions shouldn't influence the results.

3.2 Methods

First, the amount of time students spent on the site was analyzed; this was done to get an overall view of how much the students used the site. The time a student spent on the site was summed over the entire course duration; any activity, for any amount of time, was counted. Students were grouped by their final grade letter and the time spent on the site was averaged for each grade group.

Next, the number of videos watched each day was analyzed. The aim of this was to start to see if students were watching videos as they were being told to; the analysis hoped to answer if students were watching the pre-lecture videos before class, or at no particular time. The number of videos watched by all students for each day was determined and plotted in a bar graph.

Student data was analyzed according to class times to get more detail on when students were watching videos. Since there were three different times for lectures, students understandably watched the videos at various times. A simple point system was created to determine if students were watching videos on time and if that correlated with better grades. The point system works as follows: 1) if a video was watched, a student gained one point, 2) if the video was watched before the student's lecture time for the video's corresponding day, the student gained another point, and 3) if the video watched was not on the list of required or suggested videos, that student gained no points. Points were totaled separate for if the video was required or suggested, then these two were added together to determine the total score earned. Note that this point system was only used for data analysis and was created after the end of the course, so it did not in any way affect the students or their grades.

Additionally, we were interested in the students' average interaction with pre-lecture videos based on their grades; the class grade, weighted average exam grade, and individual exam grades were available. The students' interactions with the videos were measured by how many videos they watched, how much time they spent watching videos, and what percent of videos they watched in terms of the number they were expected to watch. First, the average time spent on videos for students who earned certain grades was determined. The amount of time each student was on a video page was summed and then plotted in a bar graph, ordered by class grade; this graph was not very intuitive, so simple scatterplots of time spent versus class grades or weighted average exam grades were made. Again, the scatterplots did not successfully answer the question being asked, so the students were grouped by their class letter grade and the interaction time for the group was averaged; this was plotted as a bar graph. This procedure was repeated for the number of videos students watched; this was found to be a better method because we were more concerned with if the videos were watched rather than the length of time they were watched. The average number of videos watched by students who earned each letter grade were plotted as bar graphs for their weighted average exam grades, final grades, midterm 2 grades, and class grades. The students were also grouped by the number of videos watched (0-10, 10-20 videos, etc.) and then the groups' average grades were plotted in a bar graph. The problem with this method is that the population of each group varied largely; one of the groups had 96 students while another had only one. This is a problem that occurred in most of the graphs that grouped students but the standard deviation was difficult to calculate because of all the conditional statements used in the analysis, so instead the group size is presented in each plot as needed.

The previous methods showed trends in how students interact with videos, but neither analysis accounted for how much interaction was expected of students. So, to get a better understanding of how students compared to expectation, the number of videos watched was converted to a percentage of videos watched. The number of videos watched was divided by the number of videos that were expected to be watched; this calculation did not account for if a student repeatedly watched a video, so a student could have not watched all the videos, but the analysis may show that they had if they repeatedly watched other videos. Initially, the average percent of videos watched was calculated over the entire term and then compared to grades. Then, to see how the percent watched changed with each exam, the percent of videos watched before each exam was determined.

The grading scale used in the course was not the common method used, where letter grades are spaced equally by 10%, but rather it accounts for the difficulty of the course. Table 3.1 describes the grading scale used in the introductory physics course studied; as the table shows, the letter grades have various ranges. We were interested in the effect this had on the data analysis, so the percent of videos watched before the first midterm was plotted with the letter grades, as before, as well as with grouping grades into 10% increments.

A	A-	B+	B	B-	C+	C	D	F
85-100%	80-84%	77-79%	68-76%	65-67%	62-64%	50-61%	45-50%	0-44%

Table 3.1: The grading scale used in the introductory physics (ph 201) course studied.

Finally, the gains were explored to determine if there is a correlation between watching more videos and earning a better grade. The change in scores on the exams were used to determine gains; the exam scores had to be normalized since the difficulty of each was different. The score received on each exam was divided by the average class score for each student to normalize the gains. The number of pre-lecture videos to be watched also differed between each exam; to normalize this, the number of videos watched was divided by the number of videos assigned to obtain the percent of videos watched. Gains in percent of videos watched and gains in exam scores were determined for each student; these were plotted as a scatterplot. The scatterplot was not as intuitive as was desired, so the gains were grouped into bins; for example, students who had increased their percent in videos watched between .2 to .4 percent were grouped together and their gains in exam scores were averaged, and this was repeated for other ranges of percentage gains

While studying the 2016 dataset, the research team discovered that the BoxSand website wasn't documenting when a student right-clicked and opened an external link in a new tab. About two-thirds of the pre-lecture videos used in this analysis were located on a different website, and therefore some data on video interactions is missing. To determine how much this may have affected the relationships, the trend in average number of offsite videos watched was compared to the trend in the average number of onsite videos watched.

3.3 Results

The average time that students, grouped by course grade, spent on the BoxSand website for the entire term was plotted and presented as Figure 3.1. The results show that on average, C students spent the most time on the BoxSand website, followed by A and B students. The number of students in each group is presented in the plot above each bar. The amount of time students spent on the website was averaged in hours.

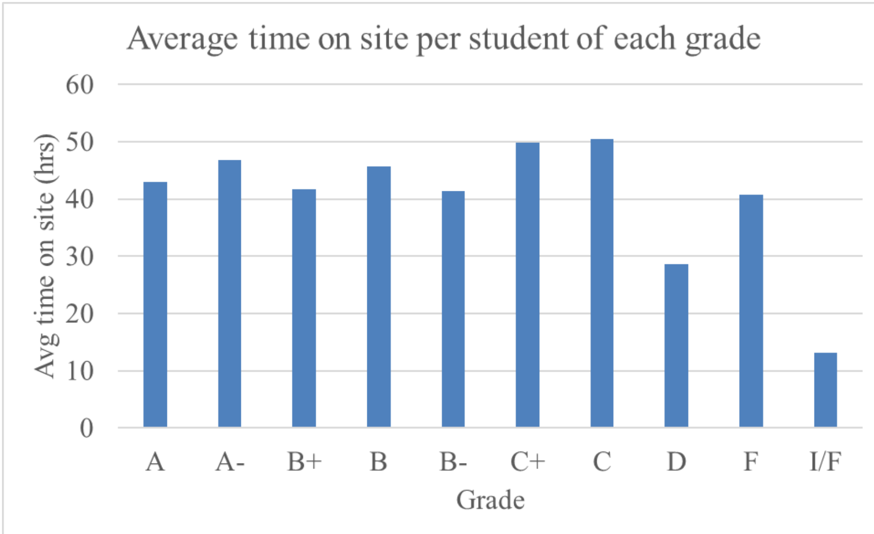

Figure 3.1: The time in hours that each student spent on the BoxSand website was calculated. The students were then grouped by letter grade (course grade) and the time spent on the site for each group was averaged.

The number of videos watched by the class each day was plotted, see Figure 3.2. The number of videos watched peaked and fell throughout each week with a relatively similar pattern. At the end of the week, Fridays and Saturdays, the number of videos watched drops off. Towards the beginning of the week, Sundays and Mondays, the number of videos watched goes up. Also, near the exams, there is an increase in the number watched with a subsequent drop off after the midterm days.

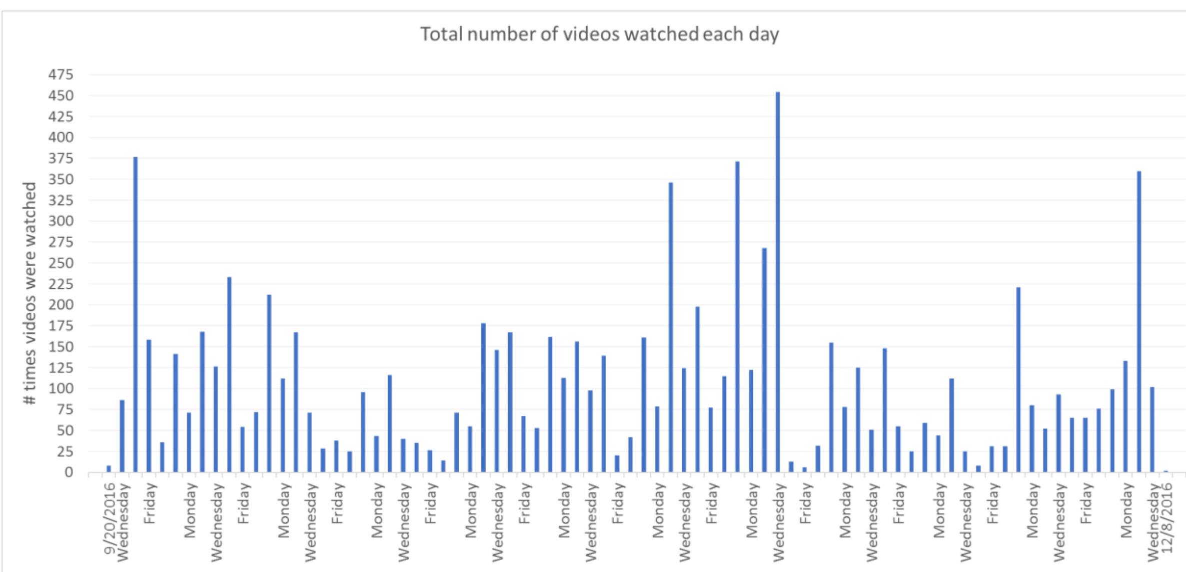

Figure 3.2: The number of times videos were watched each day throughout the term.

For a more detailed approach that factored in lecture times, a point system was created; see the methods section 3.2 for details on the point system. The more points a student has, the more videos they watched and/or the more videos they watched on time (before their lecture time). In figure 3.3, the average number of points earned by students of each weighted average exam grade is plotted in a bar graph. The group size is presented above each bar. The middle grade ranges, B to C+, had the highest average video interaction score. Students who earned an F averaged the lowest video interaction score.

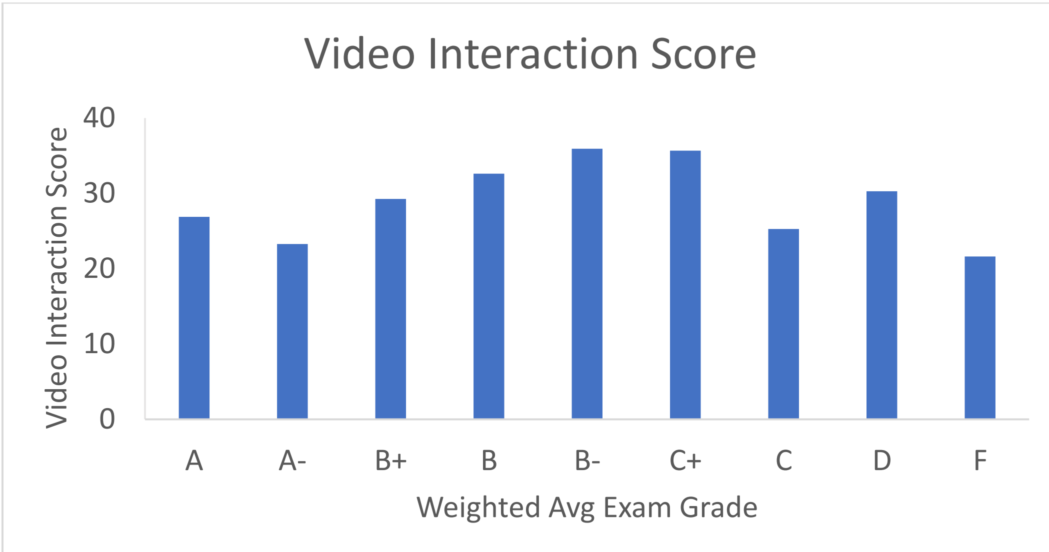

Figure 3.3: The video interaction score is a figure of merit that was determined by when a student watched a video, see section 3.2 for details. The average score is plotted for each grade group.

The two variables used to compare video interaction and grades were 1) the length of time a student spent on videos, and 2) the number of videos a student clicked on. The data analysis shifted towards analyzing how many videos were clicked, so the figures presented in this section will reflect that.

The average time that students of each grade spent on videos for the course is presented in Figure 3.4; the class grade was used and the time was measured in minutes. Students in the mid-grade range, B- to C, had on average the most time spent watching videos. Students who received an F or an incomplete also spent a larger amount of time watching videos, however the group size is small and the standard deviation is large.

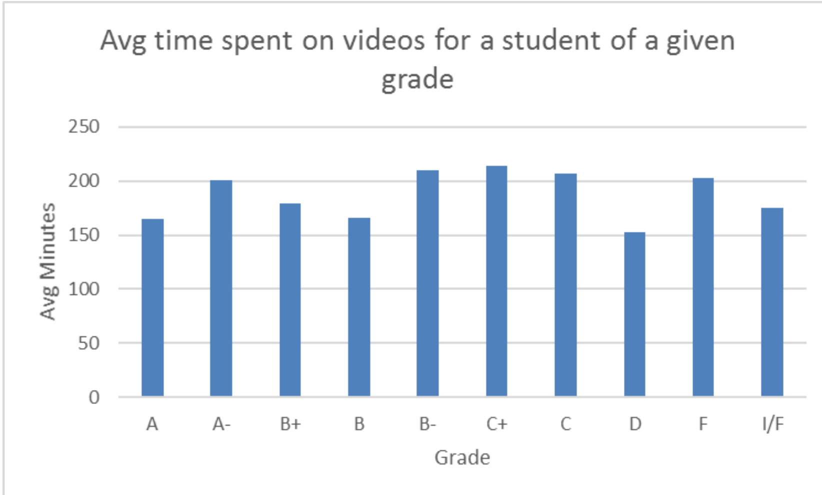

Figure 3.4: The students were grouped by course grade and the time spent on videos over the entire term (in minutes) was averaged for each group.

The average number of videos watched by each student was determined and various analyses were done to view the data. First, the students were grouped by class grade and the average interaction of each group was found and plotted, see Figure 3.5. The results are comparable to Figure 3.4 where the time watched was analyzed instead of number of views; the main difference between the two is that F students watched, on average, the largest number of videos. Next, the students were grouped by their weighted average exam scores, Figure 3.6. Similar results are noticed again, where the students who earned a midrange, B to C+, exam grade watched on average the largest number of videos; however, this time the lowest averages were of the A, C and F students.

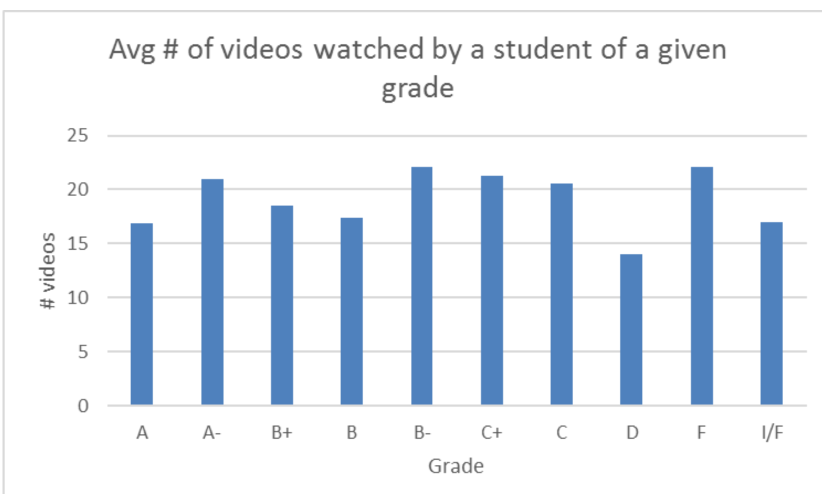

Figure: 3.5: The students were grouped by course grade and the number of videos watched by each group was averaged.

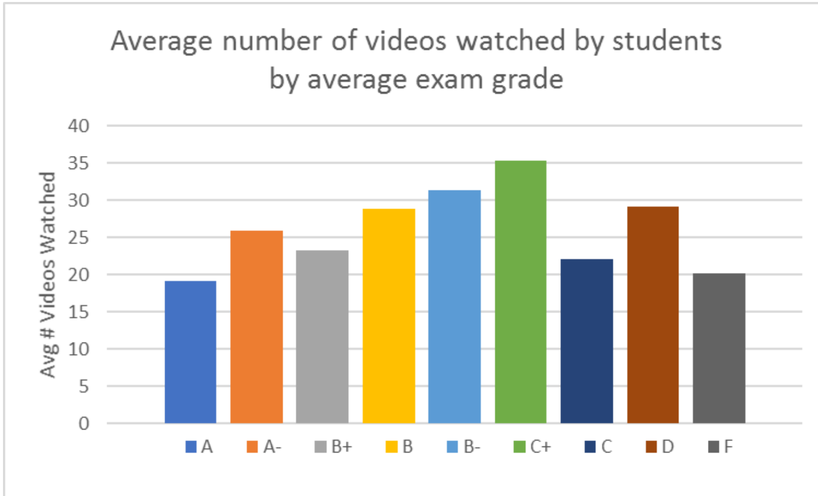

Figure 3.6: Students were grouped by their weighted average exam grade and the average number of videos watched by a student in each group was found.

To focus more on the interaction level of students, the students were grouped by how many videos they watched, and then the average grade of that group was determined; figure 3.7 explores this relationship. The groups with the highest average grades watched 30-40 videos or more than 70 videos, and the lowest average grade came from the group who watched no videos.

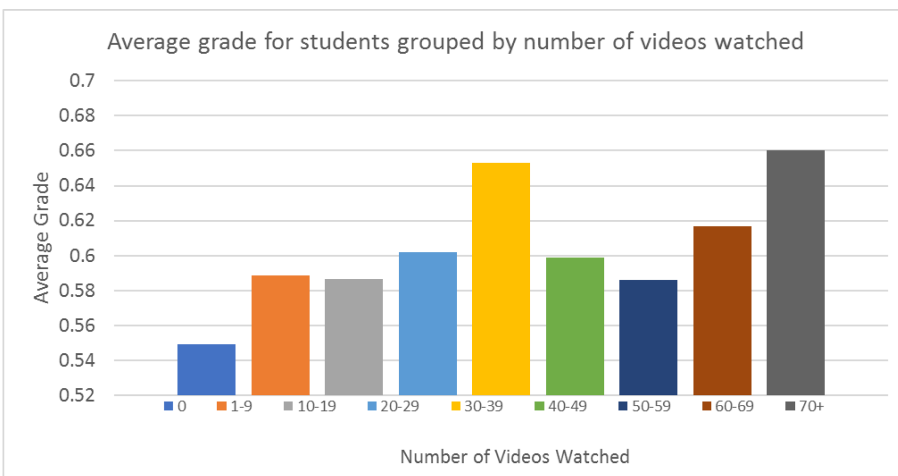

Figure 3.7: Note that the grade axes starts at 52% and goes to 70%. Students were grouped by the number of videos they watched for the term. The average grade (weighted average exam grade) for each group was found.

Looking at the number of videos students watched provides a way to observe trends, but it is somewhat abstract because it doesn't provide information on the number of videos that were expected to be watched. The number of videos assigned to be watched before each exam differed between exams, so the average percent of videos watched was determined based on the videos watched before each exam and on each student's grade on that exam; the results are plotted in Figure 3.8. The initial observations to make from Figure 3.8 are on average, students 1) watched less than 50% of the assigned videos, 2) watched a larger percent of the required videos than the suggested videos. Another observation one can make from Figure 3.8 is that a higher percent of videos was watched between the midterms than for before the first midterm or after the second midterm; this could be for two reasons 1) students were watching more videos, or 2) more videos were trackable for the middle third videos because they were located on the BoxSand website and even the videos opened in a new tab or window were trackable (while they weren't trackable during the other parts of the term).

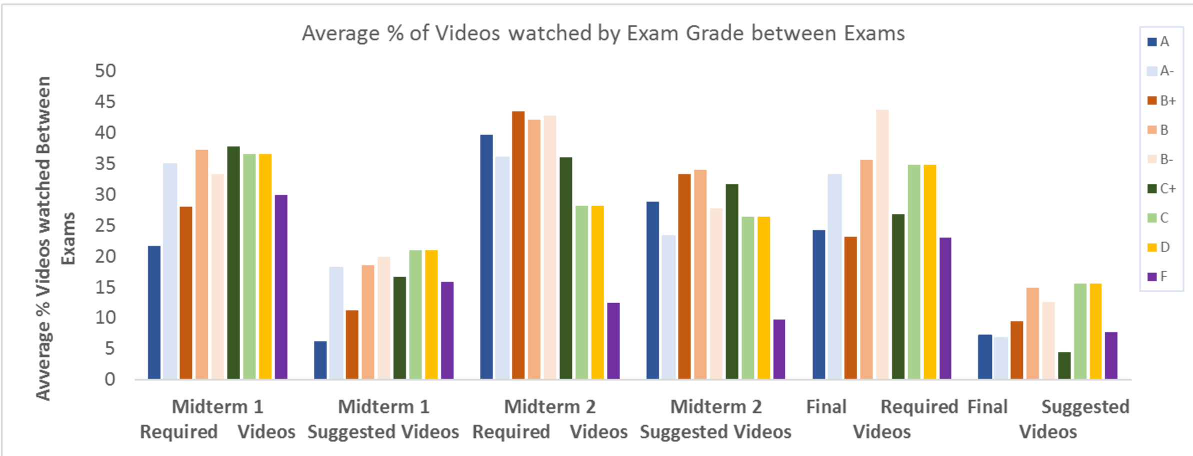

Figure 3.8: Students were grouped by their exam score and the average percent of videos watched by the group between each exam was determined and plotted with the students' exam score for each corresponding exam. An example of how to read the graph is: "a student who earned an A on the second midterm watched about 40% of the required videos and about 30% of the suggested videos that were assigned after the first midterm and before the second midterm".

The average number of offsite videos watched was compared to the average number of onsite videos watched to determine how not being able to track all the right-clicks to offsite videos affects our analyses. The trends of watching the onsite videos and of watching the offsite videos, after being scaled, were very similar, see Figure 3.9. There were about twice as many offsite videos as onsite videos (videos located on the BoxSand website), which is reflected in the axes scales in the plots of Figure 3.9.

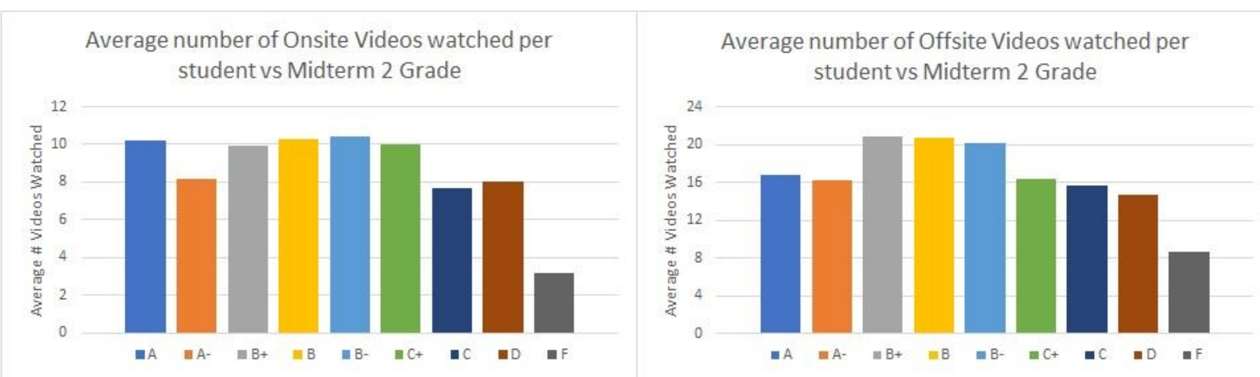

Figure 3.9: The average number of onsite and offsite videos watched per student in each grade group was determined. The side-by-side analysis is used to show how not being able to track videos opened in a new window or tab may have affected data analysis.

The grading scale used in the course was not the common method used, where letter grades are spaced equally by 10%, but rather it accounts for the difficulty of the course. The percent of videos watched before the first midterm was plotted with the letter grade (Figure 3.10) as well as with grouping grades into 10% increments (Figure 3.11). The trends in the two plots are similar; the students with the highest grades, on average, watched fewer videos and the students with a midrange grade have the highest averages. The purpose of making these plots was to determine if using the grading scale used in the course hid details about trends.

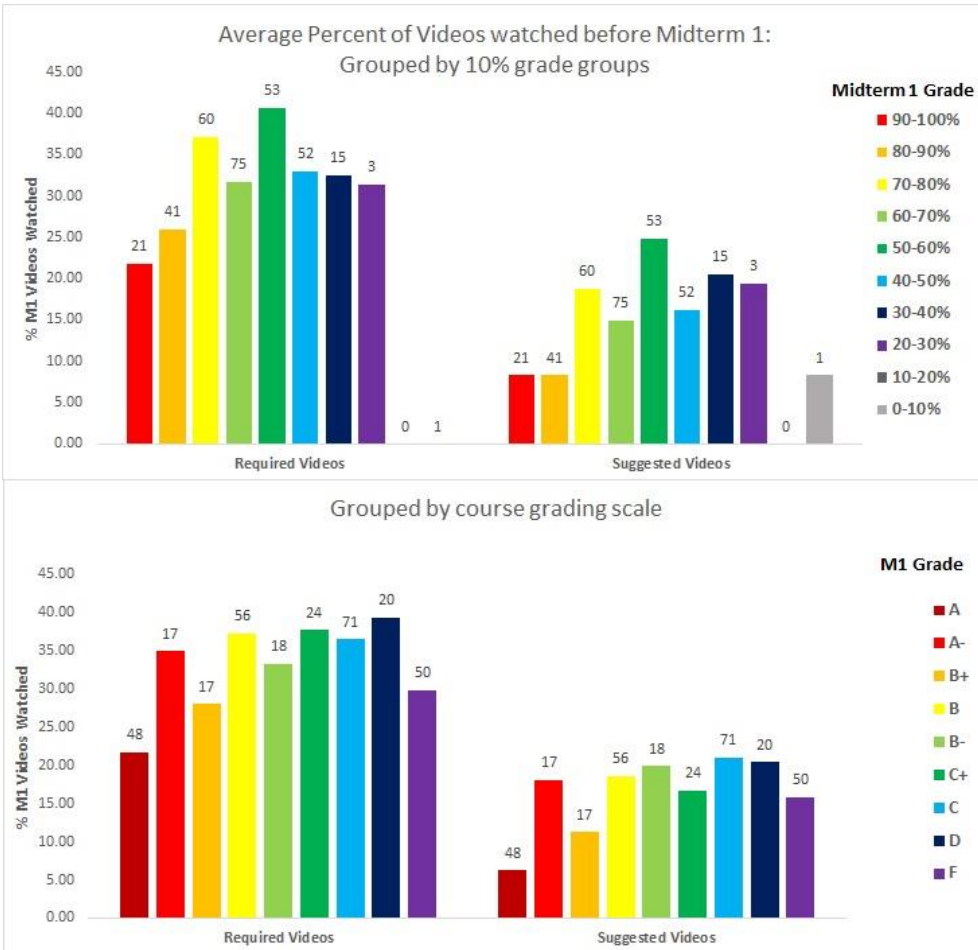

Figure 3.10: Students were grouped by their midterm 1 grades, with groups ranging by 10% (top) or by the letter grade corresponding to the course grading scale (bottom). The average percent of videos watched by the group before the midterm was determined for both required and suggested videos.

Between midterms 1 and 2, the gain in percent of videos watched and the gain in exam grade were calculated for each student. The two trends that were expected are the first two bars on the left of Figure 3.11; the majority of students followed these trends.

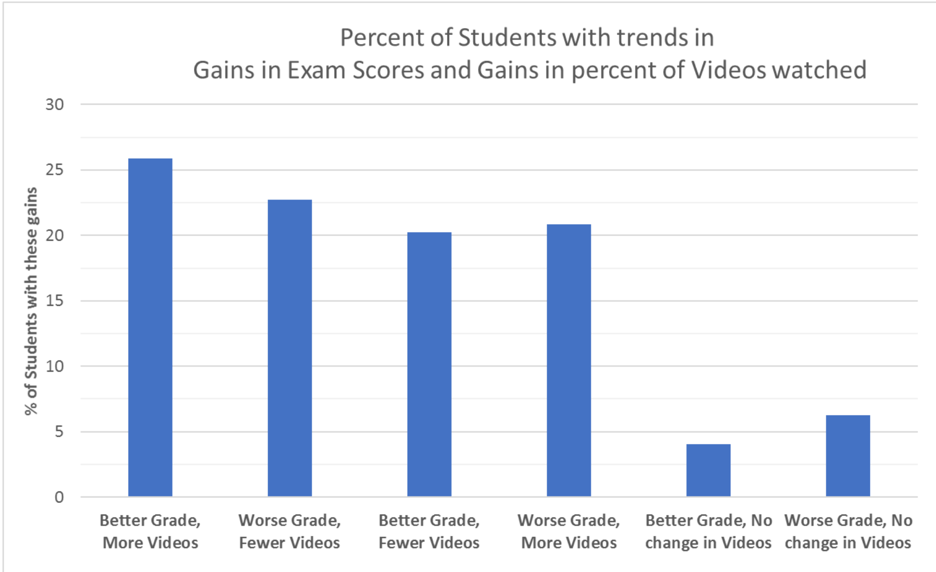

Figure 3.11: This plot counts the number of students that exhibited each trend in their gains in videos watched and gain in grade going from midterm 1 to midterm 2.

As mentioned in the methods section, many graphs were created that were not useful for analysis. These graphs are included in Appendix A. The extra plots are not crucial for understanding the process of this project, but they do show some of the steps involved to get to the desired analyses.

3.4 Analysis

We expected to see students who earned an A or B grade to have spent more time on the BoxSand website. The results show that, on average, students who earn a C spent more time on the website. Since ph201 is an introductory course, it is likely that some students had previously taken a course that covered the material, perhaps in high school. Students who have already learned the material before may not have needed to study as much, so they may not spend as much time on the site and as we see in Figure 3.1, the students with higher grades tend to use the website less. The standard deviation for the F students is very large because there were only a few students who received this grade, this helps to explain why the average time spent on the site for this group is higher than expected. Students who earn a C grade tend to need more help, so a higher average time spent on the site is expected. Further research will analyze the same students in the next courses (ph 202 and 203) and we expect to see that students will continually need to increase their interaction with course content to earn higher grades since the material is likely to be new to them.

It was assumed that pre-lecture videos were watched the day before or the day of a class. Since the classes were at different times of the day (8am,9am, and 3pm), the distribution was not as clear cut as it may have been with only one class time accounted for. As Figure 3.2 shows, there are peaks of activity before a class day with the highest instances being towards the middle of the weeks. Students seem to have been watching the pre-lecture videos to help them study/prepare for exams; there was an increase in the number of videos watched before the exams, with a drop off after the exams. This was an expected trend since watching videos is a lowkey way to study, unless students are interacting with the video; interacting with the video includes taking notes while watching it and pausing and replaying parts of the video that were not initially understood. Another anticipated trend was that students watched more videos at the beginning of the week (Sunday and Monday), and watched fewer videos at the end of the week (Friday and Saturday).

It was hoped that students generally watched videos before the corresponding lecture, but this was not necessarily expected. Table 3.2 breaks down in the simplest model what the interaction score implies; the table does not account for watching a video more than once. The video interaction score attempted to determine if students were watching videos before the lecture and how that correlated with grades. The average score for each grade group was less than 50, which indicates that on average, students don't watch all of the pre-lecture videos assigned. The middle range of grade groups (B to C+) and the D students had the highest interaction scores. This result can be interpreted in many ways, one being that these students are studying more to get their grade to the next grade level, for example D students want to keep their grades up so they don't fail the course.

Video Interaction Score		
Watched:	Before Corresponding Lecture	After Corresponding Lecture
All the Required Videos (50 total)	100 (2 points per video)	50 (1 point per video)

Table 3.2: The video interaction score is a figure of merit that indicates how often students are watching pre-lecture videos before the corresponding lecture. This table provides simplified, expected values depending on when students are watching videos.

The average time students in each grade group spent watching videos gave interesting results (Figure 3.4). The A, B, and D students had the lowest average time spent on videos. This could indicate that students with these grades are confident in their grades or understanding of the content and don't feel like they need to watch as many videos. The students who on average spent the most time watching videos may be interacting more with lecture material to try to get to the next higher grade as these students are in the grade groups of +/- a letter grade.

The plot of average number of videos watched by students in each grade group (Figure 3.5) has very similar trends as Figure 3.4 where the average time spent on videos was analyzed. This makes sense, it is expected that a strong correlation exists between these two plots, since watching more videos equates with spending more time on videos. The difference between the time and number of videos watched likely comes from students watching videos for more or less time than the length of the video. The most noticeable difference in Figures 3.4 and 3.5 is among the B- and F students who appear to watch more videos, but for less time than the other grade groups.

Figure 3.6 seems to show the opposite trend we expect and would want to show students. The figure seems to indicate that more videos watched correlates with worse average exam grades. Instead, this figure could be used to show that just watching pre-lecture videos is not enough for student to have a good understanding of the content. Analysis of other study material or lecture content could improve this message and is the topic of future research with this project.

There were 50 pre-lecture videos assigned during the course. Figure 3.7 was created to further examine the correlation between the number of videos watched and average exam grade. The figure shows that students who watched 30 to 39 or 70 or more videos on average earned the highest grades on the exams.

As we would expect, the students who didn't watch any of the videos earned the lowest exam grades. Surprisingly, the students who watched 50 to 59 videos had among the lowest average grades. Students who watched all of the videos assigned were not necessarily interacting with the videos in a way that allowed them to internalize the information. This plot suggests that how the students interact with the content might matter more than how much they interact with it.

The average percent of videos watched for each third of the term by each grade group is presented in Figure 3.8, which shows the results for both required and suggested videos. Students were not expected to watch the suggested videos, so it makes sense that the average student watched a smaller percent of suggested videos than required videos. The percent watched for the second midterm was higher than for the other exams. The middle third videos were located on the BoxSand website, so even if someone opened the video in a new tab or window, the site tracked them, but the other two thirds of the term contained offsite videos that were inconsistently tracked. One trend that the figure shows that the average A students watched a higher percentage of videos before the second midterm when comparing their interaction with B and C students and D students had the opposite trend. Various conclusions can be drawn from this, but the most interesting may be that this is a result of tracking right-clickers. Although it is unrelated to the purpose of the project, this analysis poses questions about how grades and opening pages in new windows or tabs are correlated. Figure 3.8 seems to show that A students tend to open pages in new windows or tabs, while D students tend to open the page in the same tab; however, this ignores the change in content presented in each third of the term.

Not being able to track right-clicks to offsite videos did appear to have a small impact on the observed trends when averaging groups of students' data in Figure 3.9. The most noticeable trend comes from A students who averaged among the highest number of videos watched when the videos were trackable (onsite) but not when the videos were offsite and some video interaction was not trackable.

The grading scale used in the course accounts for the difficulty of the course, so the groups do not have an even amount of percentages in each. To test the effect of this on the analysis, the course grading scale was compared to an evenly spaced grading scale that groups students by 10% grade spacings, see Figure 3.10. The plot that used the course grading scale had a more even spread of students in each grade group than that for the evenly spaced by percent grade groups. The plots show similar overall trends, so it is assumed that either grading scale works well for analysis and the one used for the course will be used for the rest of the data analysis.

The gains in exam grade versus gains in percent of videos watched had six possible trends that are equivalent to counting the number of dots in each quadrant and on each axis (where there was no change in videos watched) of a scatterplot of the gains. The desired trends were the first two on the graph and had the largest percent of students in them: watching more videos correlates with better grades, or watching fewer videos correlates with fewer videos. The students who showed no change in the percent of videos watched implies that they are consistent in how they watch videos and that after the first midterm they did not make an effort to improve their video watching habits.

3.5 Summary

The Fall 2016 dataset was analyzed to find correlations between video interaction and grades and to get an overall idea of how students used the BoxSand website. Watching more videos tends to correlate with better grades and students who increased the percent of videos they watched prior to an exam had the best chance of improving their grades on the next exam. Information about the videos that were not located on the BoxSand website was not being recorded if the links to the videos were opened in a new tab or window. The analysis indicated that this had a possible effect on the data trends.

4. Conclusions

The BoxSand website provides us with important information regarding the behavior of students when interacting with course material. This study explored student interaction with pre-lecture videos via the BoxSand website and some important observations were made. Introductory physics students may benefit from watching pre-lecture videos using the BoxSand website; individual cases cannot be predicted, but there is a general trend that shows watching the pre-lecture videos correlates with better grades. However, the results also indicate that improving interaction with videos does not correlate strongly with improved exam scores; this is an important observation that shows students the necessity of not relying on one type of instruction for studying and to form a habit of studying using more hands-on examples. The students were found to usually stay on a video's page for the duration of the video, indicating that they were watching the entire video. Much of the students were not watching pre-lecture videos before class, indicating that the instructor could place more emphasis on watching them on time, this is an example of how the website can be used to improve instruction.

Future work with the BoxSand website will aim to find the best methods for teaching introductory physics and to be able to predict a student's grade based on engagement with course content. This will be done by observing how interaction with different types of content influence grades, and assumingly learning. A prediction model created from these results would only show the average or usual results seen, we cannot prove that one style of learning is definitively better than another. Future research will be able to factor the various types of learning content together to get a better model.

5. References

1. R. Baker & P. Inventado; *Chapter X: Educational Data Mining and Learning Analytics;* pp.1,3
2. A. Merceron & K. Yacef; *Educational Data Mining*: a *Case Study; pp. 1, 7-8. (2005).*
3. Calvet Linan, Laura; Juan Perez, Ángel Alejandro. *Educational Data Mining and Learning Analytics: differences, similarities, and time evolution.* RUSC. Universities and Knowledge Society Journal, [S.l.], p. 98-112, jul. 2015. ISSN 1698-580X

Appendix A

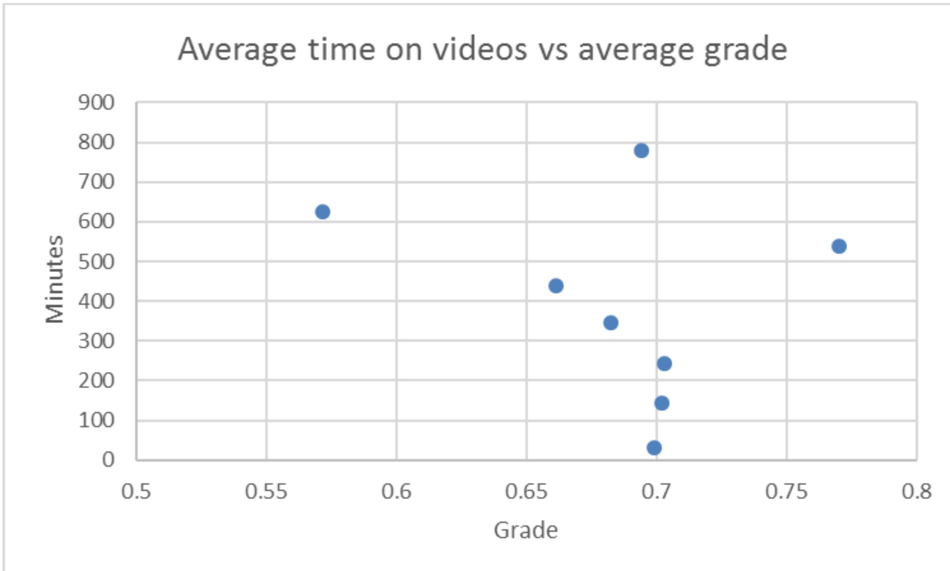

Table A.1: Sample sizes for grouping students by engagement. Time watching videos is in minutes.

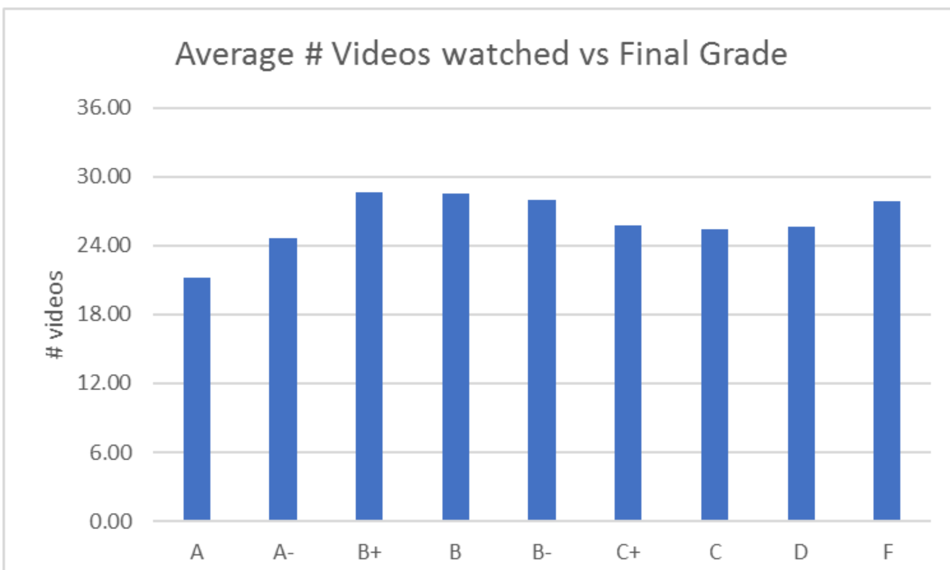

Figure A.2: The average number of videos watched by each grade group using the course grade.

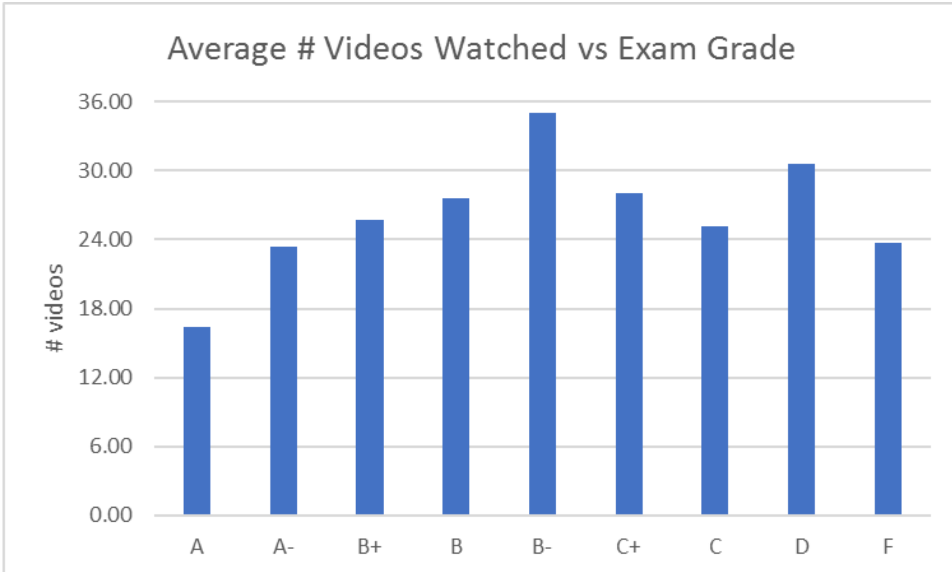

Figure A.3: The average number of videos watched by each grade group using the course grade.

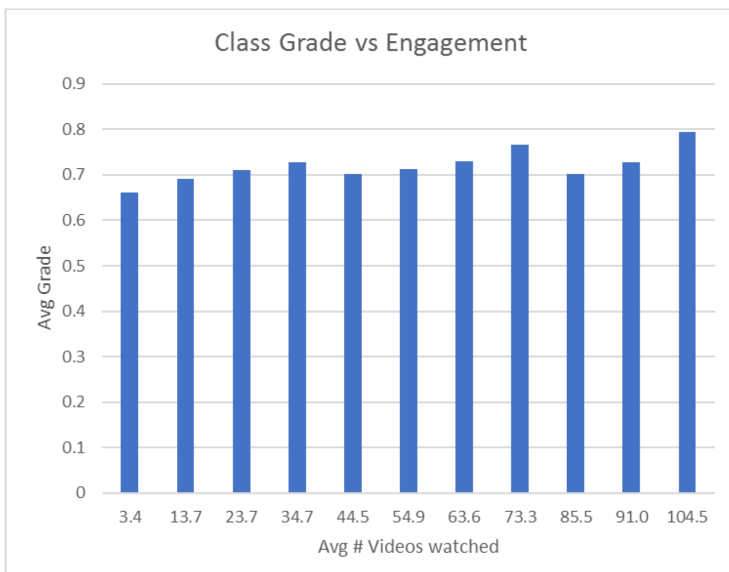

Figure A.4: Looking at how the number of videos watched affects grades.

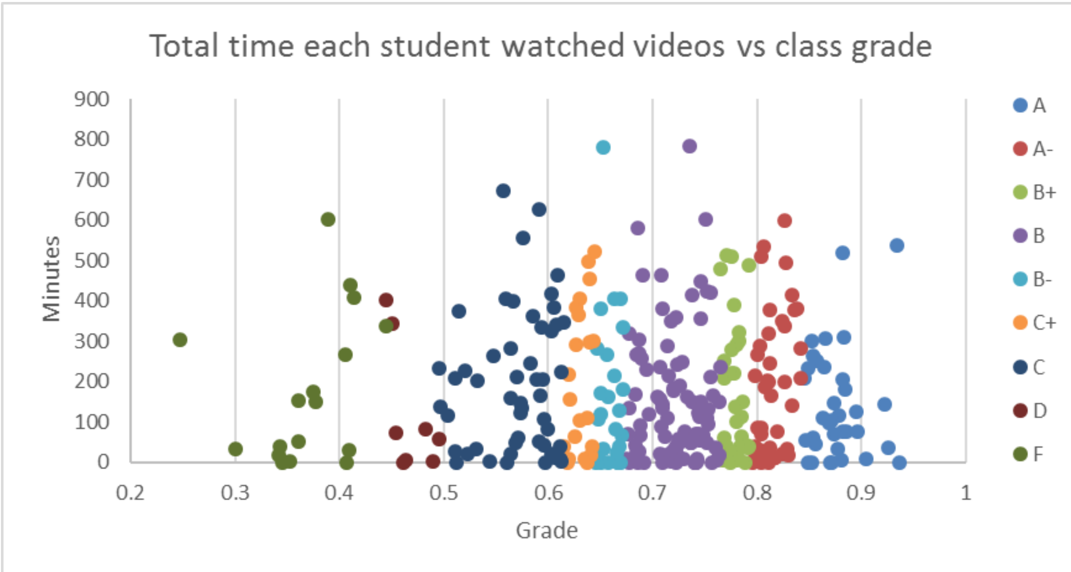

Figure A.5: The time a student spent on any page containing a video was totaled in minutes. The total was graphed versus the student's final class grade

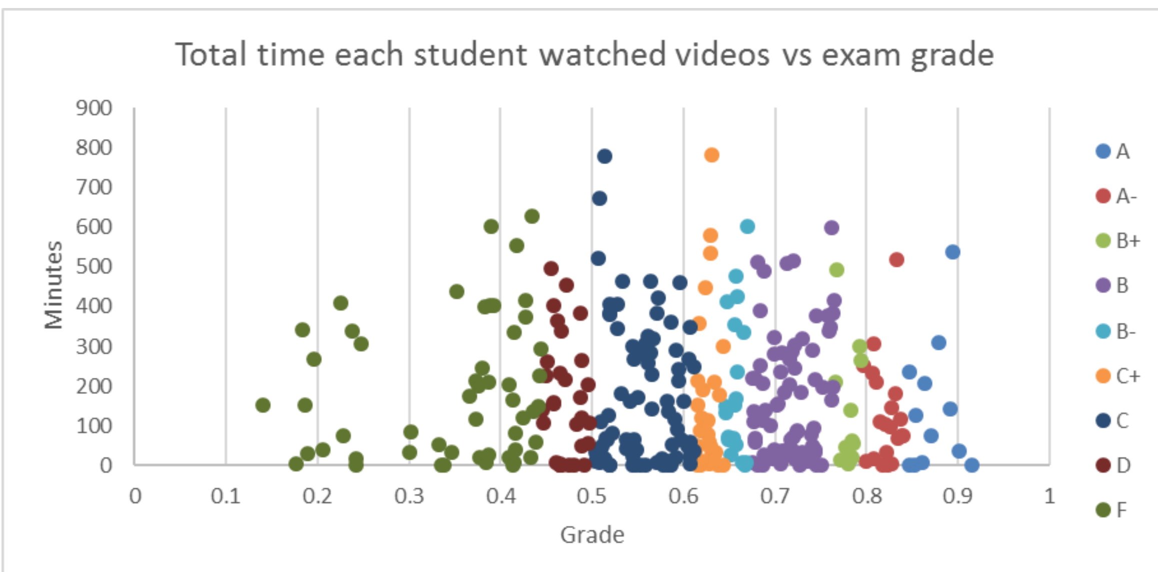

Figure A.6: The time a student spent on any page containing a video was totaled in minutes. The total was graphed versus the student's exam grade.

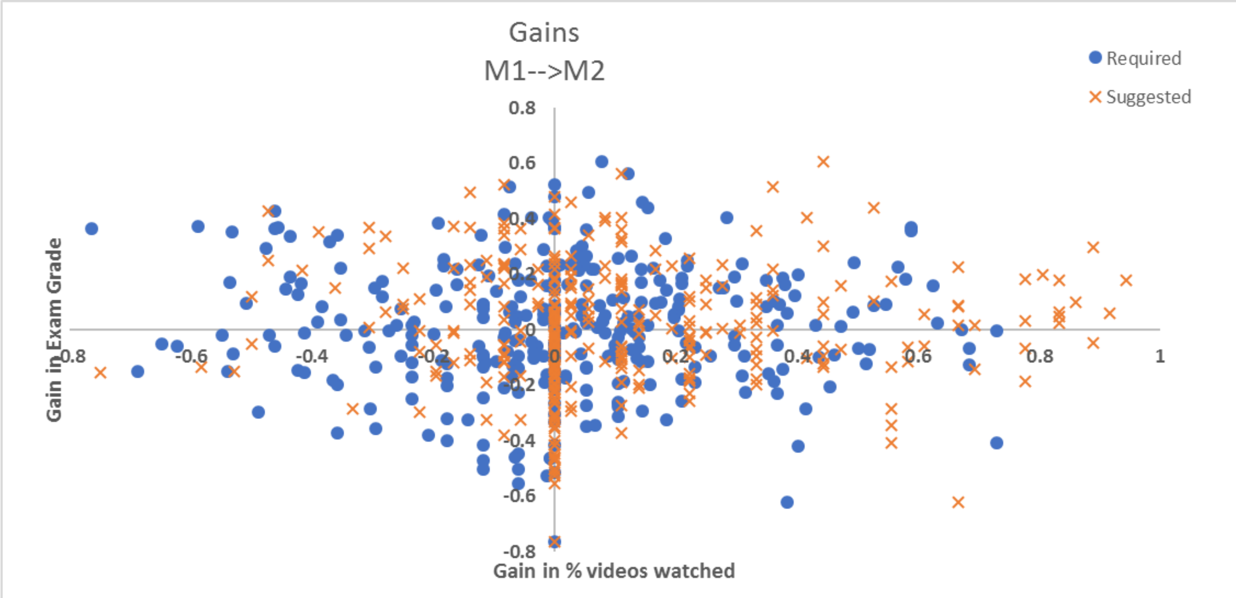

Figure A.7: Gains in grade versus gains in percent of videos watched between Midterm 1 and Midterm 2.

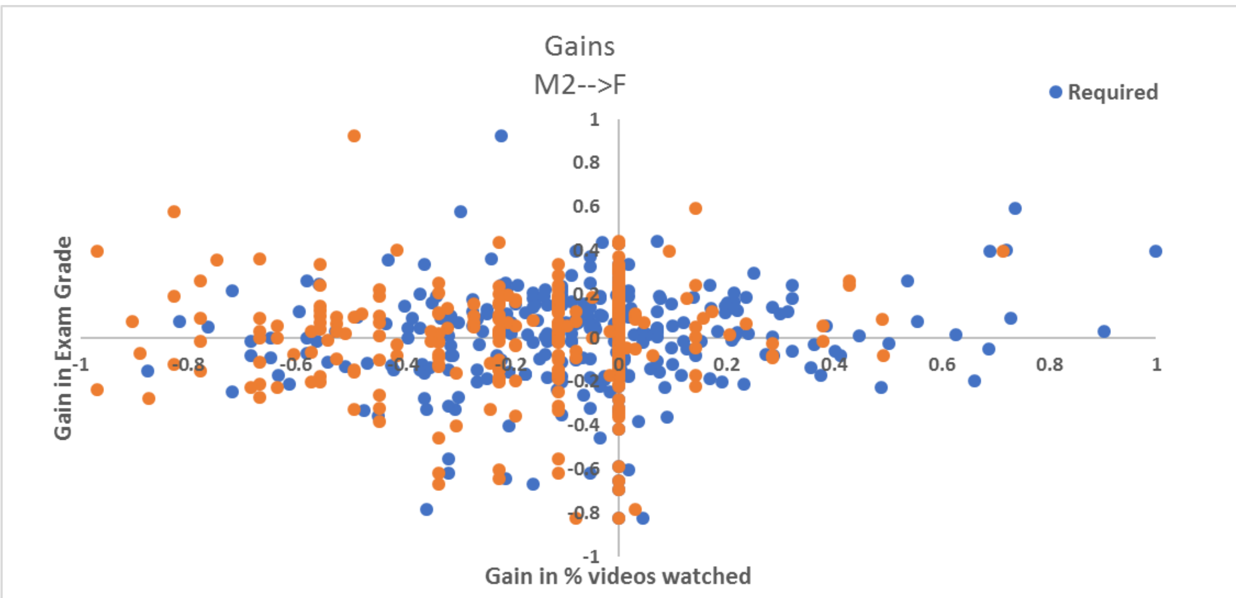

Figure A.8: Gains in grade versus gains in percent of videos watched between Midterm 2 and the Final. The orange dots represent suggested videos.