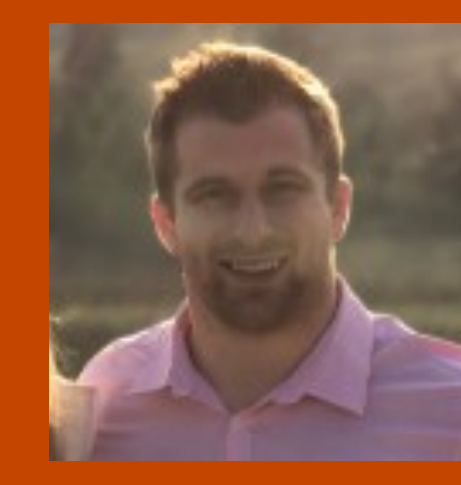




K.C. Walsh
OSU Physics
boxsand.org/walsh



Michael Dumelle
OSU Statistics



Katy Williams
OSU Mathematics



Benny
OSU Beaver

Background

Our intro sequence moved away from lecture-based instruction towards an actively engaged, flipped classroom model. Roughly 300 pre-lecture videos were created along with corresponding pre and post lecture online homework assignments. Students engagement with those resources is tracked with the motivation of understanding the effects of pedagogical changes from the lens of data analytics.

Abstract

As education uses online and digital learning tools and resources more, an opportunity arises to study students' learning behaviors and outcomes through data analytics. In this study we perform correlation data mining of individual student's click-stream on both an Open Educational Resource site, BoxSand.org, and online homework on Mastering Physics. Exploratory analysis can be used to inform a model-based approach with long-term goals of creating inferential and predictive models

Study Info

- Introductory algebra-based physics, Fall 2016 - Spring 2109
- ~350 Students per year consent to tracking, nearly 70% of the class
- ~2.7 million data points on student's interaction with open resources and homework each year
- Exploratory analysis aimed to inform model-based analysis

Project BoxSand

BoxSand is an open educational resource (OER) website with links to: videos, practice problems, open source textbooks, educational websites, simulations, etc. The site is setup as a repository of custom content produced at OSU and the best OER around the web. Students are guided through the site and their out-of-class activities with the course's *Daily Learning Guide*. The BoxSand site tracks students' click-stream while they progress through their out-of-class study and practice. This information is used to understand what resources and engagement behaviors correlate to learning and to create predictive models for individualized learning paths.



Methods

- Study participants are OSU students
- Data is de-identified by a 3rd party to preserve students anonymity
- BoxSand.org was built by COSINE and OSU students using Drupal 7
- Pearson provided click-stream data from Masteringphysics
- Correlation Data Mining compares engagement with the site and online homework to grades in the course
- Exploratory analysis can be used to inform model-based predictions
- Linear mixed model analysis helps to understand effect size
- Future predictive models can be used for early detection and interventions

Correlation matrix

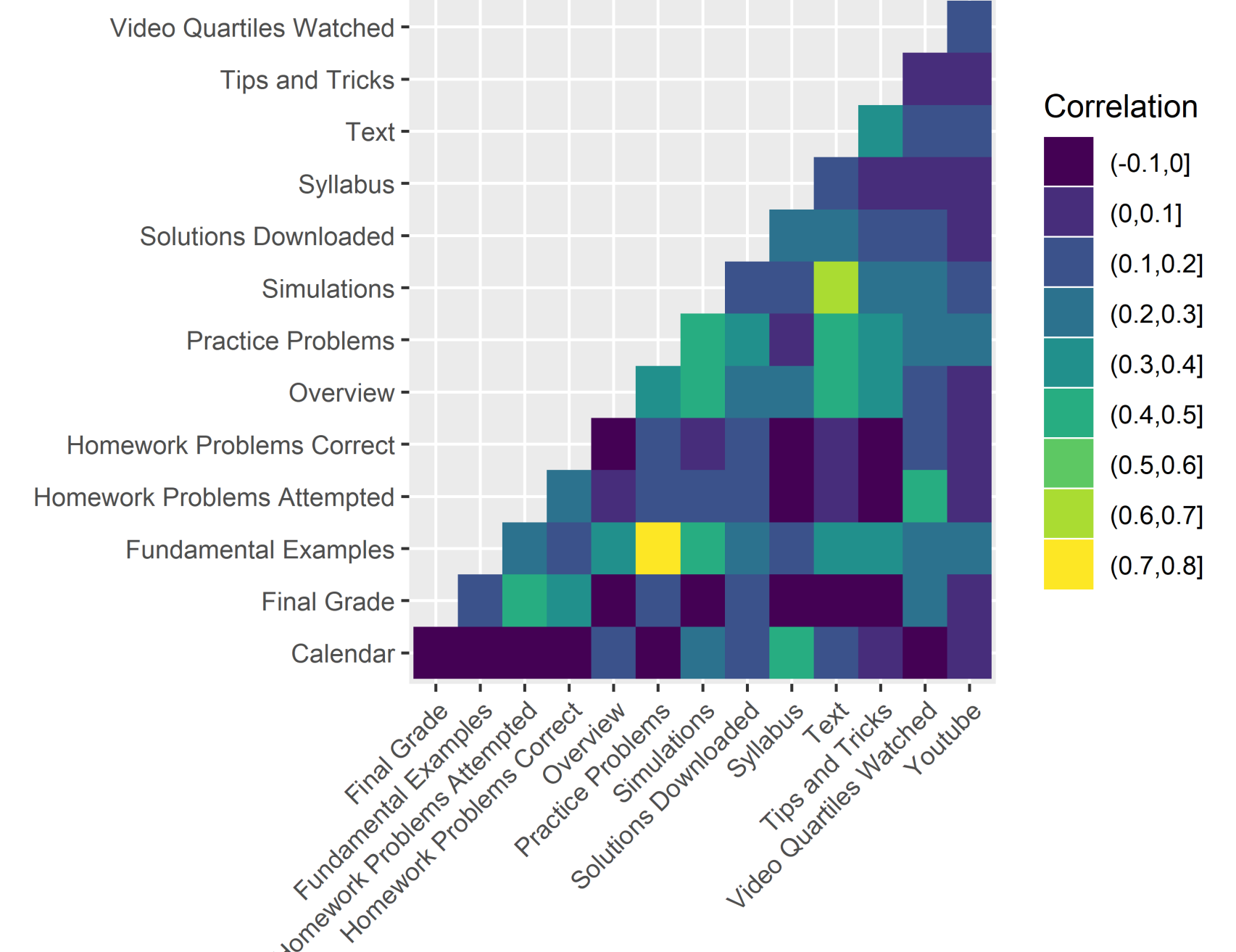


FIG 1. Correlation between several BoxSand click-stream data sources, online homework, and grades.

Our first question was to see how engaging in one type of content correlates with other types, along with grades. The correlation matrix shows the strength of the correlation for what we believed to be the 12 most interesting data sources along with final grade. The data sources above are:

- *Video Quartiles* | how much OSU pre-lecture videos watched
- *Tips and Tricks* | topic specific problem solving advice
- *Text* | OpenStax, Hyperphysics, and other text OER
- *Syllabus* | course information
- *Solutions Downloaded* | challenge homework and exams
- *Simulations* | PhET, Fendt, etc.
- *Practice Problems* | non-graded practice problems
- *Overview* | topic overviews and learning objectives
- *Homework Correct* | answering MP correctly
- *Homework Attempted* | just attempting MP questions
- *Fundamental Examples* | worked examples
- *Final Grade* | Overall course grade
- *Calendar* | course schedule including a daily learning guide

The three strongest correlators with final grade are pre-lecture video watching, online homework attempted, and online homework correct. Other groups and patterns emerge and can be used as a basis for classifying students in future predictive modeling work.

Cramming behavior

We looked at a number of variables and how they correlated with grades. Doing so on a weekly basis shows that during non-exam weeks engaging more with OER, like pre-lecture videos, is typically positively correlated with grades. This correlation is likely to flip (negative slope) during exam weeks.

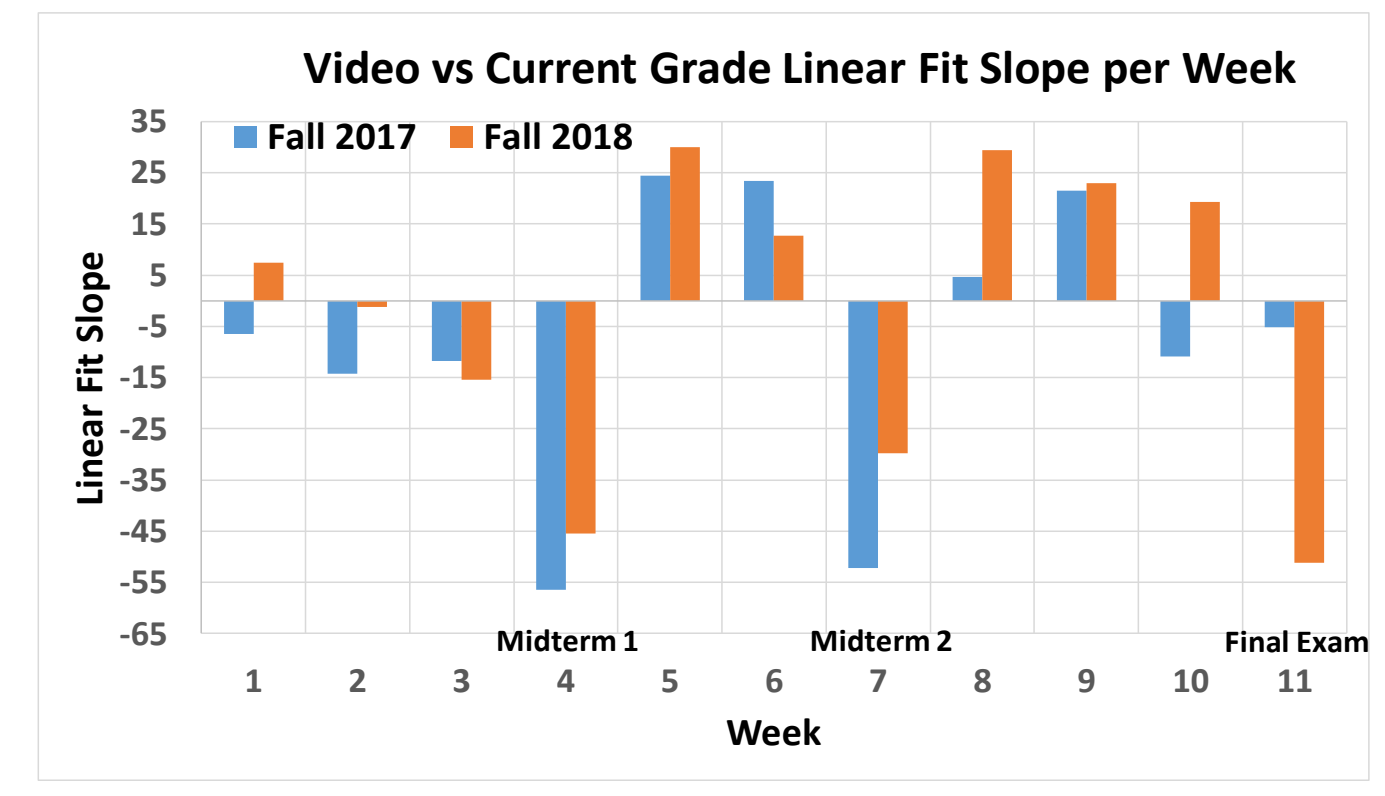


Fig 3. Videos watched vs. current grade linear fit slope in units of videos vs. 100% grade difference.

Video watching vs. exam grade

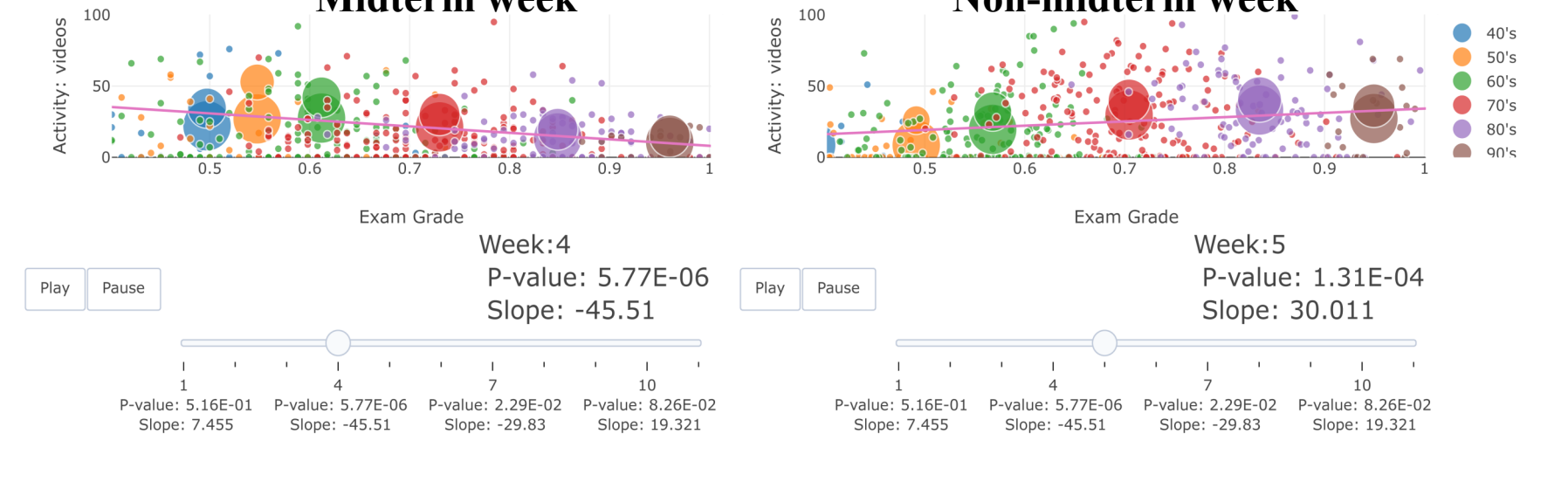


Fig 2. Videos watched vs. exam grade, week 4 (midterm week) and week 5.

We believe this is due to the cramming effect and healthy study habits. Students should be watching videos, reading text, and introducing themselves to material early in the learning cycle. When exam study comes around they should have moved onto practicing and applying physics. This is somewhat confirmed when looking at online homework engagement which does not go through such a correlation flip on exam weeks.

Online homework engagement

To explore the importance of online homework we wanted to visualize the distribution of the students engagement. The 3-D plot shows the percentage of Masteringphysics engagement on the horizontal axis vs. the % of students in each range on the vertical axis, for a given grouping of course grade. Front and to the left is low course grade and low MP engagement while back and to the right is high grade and high MP engagement. It's clear that the students performing better in the class tend to engagement more with their online homework.

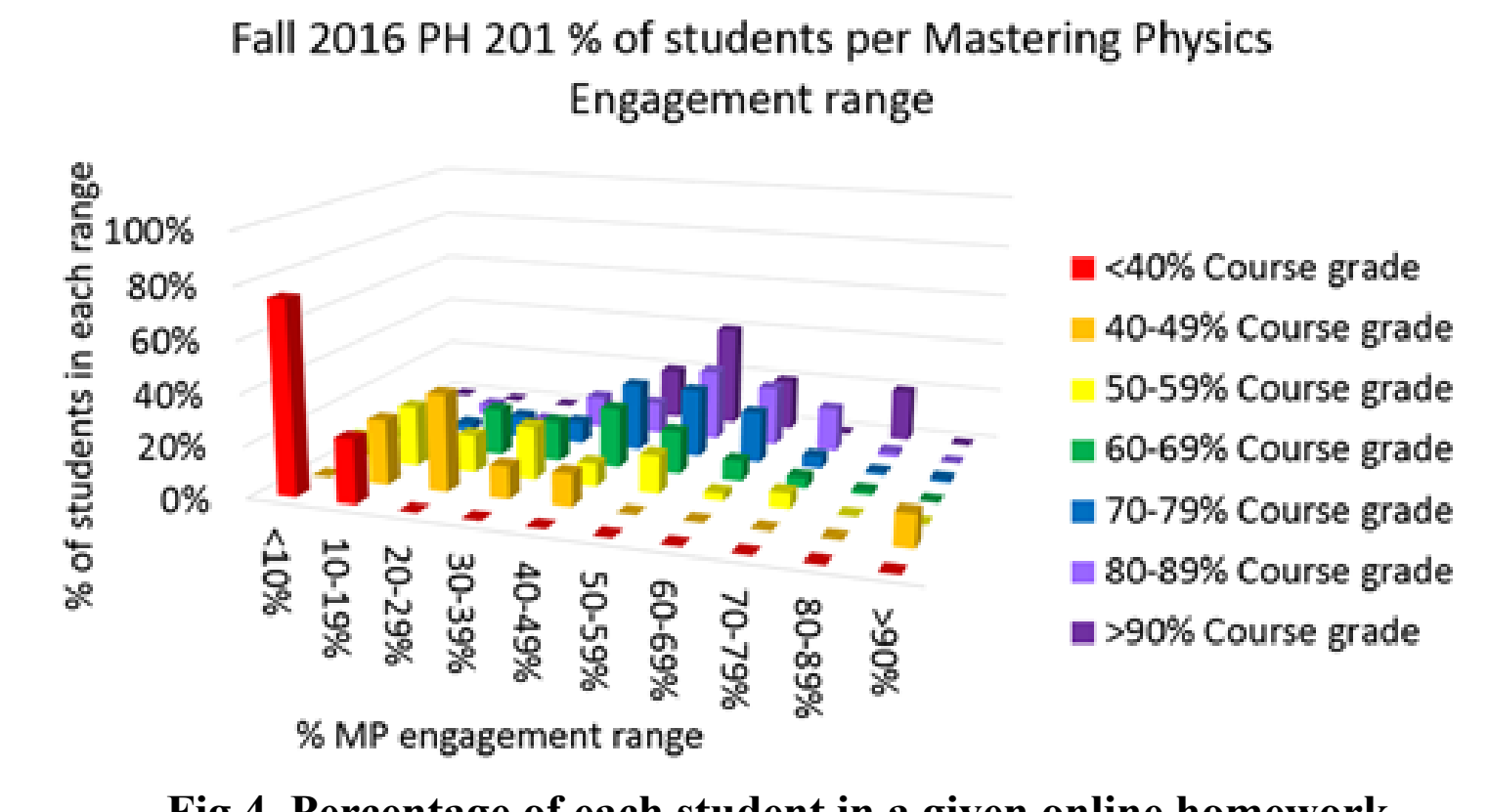


Fig 4. Percentage of each student in a given online homework engagement range vs. course grade.

Model-based Analysis

Our hypothesis was that we could create a model to predict part of the students' grade based on watching pre-lecture videos, attempting homework, and answering homework correctly. We used a linear mixed model analysis of the form: $y = \beta_0 + \beta_1 v + \beta_2 P + \beta_3 C + \beta_4 W + \beta_5 S + \tau + \epsilon$. Here β_0 represents the expected grade intercept watching no videos or attempting any homework. $\beta_1, \beta_2, \beta_3$, correspond with the expected grade increase for every 10% increase in engagement with pre-lecture videos, homework attempted, and homework correct. β_4 and β_5 correspond to the expected grade intercept shifts between winter and spring respectively. Lastly, τ and ϵ correspond with the between student and between term errors respectively. We found that our hypothesis was correct and videos and homework could be used to quantify a predictive effect towards a students grade. Videos watched can have an expected maximum effect of 2.5% on a students grade. Perhaps more surprising is just attempting homework can have a maximum effect of 13.5% on a students final grade, 2% higher than whether they answered the questions correctly or not. Being that online homework has a ~5% contribution to your overall grade, merely attempting it has a nearly 3-fold return on investment. We suspect that while getting homework correct is a significant predictor, it is being represented as less than just attempting due to the prevalence of online cheating. Either way the conclusion is clear, doing your online homework is important. We also found that the variability between terms for a given student is much less than the variability between students, which was not surprising.

Variable	Estimate	P-value
Intercept (β_0)	58.76 %	< 0.0001
Quartiles (β_1)	0.25 %	0.0044
HW Attempted (β_2)	1.35 %	< 0.0001
HW Correct (β_3)	1.15 %	0.0006
Winter term (β_4)	-6.17 %	< 0.0001
Spring term (β_5)	-9.17 %	< 0.0001

Fig 5. Table of linear mixed model coefficients.

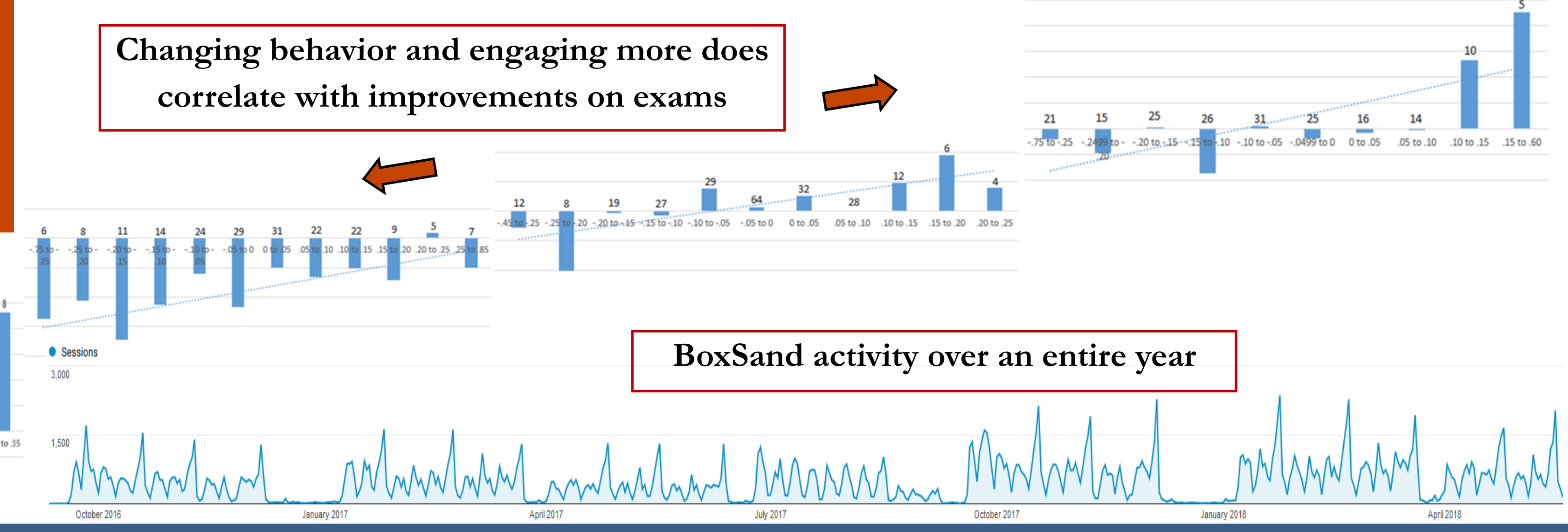
Conclusions

- Watching pre-lecture videos, attempting online homework, and correctly answering online homework is positively correlated with final grades
- Leading up to exams students should practice problems and not watch videos
- Students that increase their engagement with videos can expect to see increases on exam scores
- A linear mixed model can predict videos and homework effect on grades
- Video watching has an expected maximum effect of 2.5% on grades
- Attempting homework has an expected maximum effect of 13.5% on grades
- Answering homework correctly has a 2.5% smaller effect than attempting

Future Work

- Distance learning comparative study
- Predictive modeling using artificial intelligence
- Early warning detection with careful interventions for failing students

Changing behavior and engaging more does correlate with improvements on exams



BoxSand activity over an entire year

We are grateful to the individuals who participated in the studies associated with this work and the people who supported this work with their time and help

- Creation of the BoxSand website was made possible by funds from an OSU Learning Innovation Grant, an Open Educational Resources Grant, and support from the physics department
- Early support for the research was funded through an OSU Action Research Fellowship. Further research support was provided by an OSU Ecampus Research Fellowship
- Special thanks to the COSINE helpdesk for technical support
- Special thanks to OSU students Ikaika Mckeague-Mcfadden, Jake Bigelow, and James May

Acknowledgements