Project BoxSand: Tracking Student Engagement with Open Educational Resources and Homework

Kenneth Walsh
Oregon State University

This project was funded by the **Open Educational Resources Unit** at Oregon State University Ecampus.

Oregon State University

Introduction

I am a physics education researcher. My 'laboratory' is a year-long Introductory Algebra-based Physics sequence taught at Oregon State University (OSU). While not technically a hybrid class, it has most of the aspects of a hybrid learning environment, as the course is "flipped," and includes a combination of pre-lecture online content, followed by in-class learning experiences, as well as post-class homework. The sequence consists of undergraduate science majors, mostly in their Junior and Senior year. We collect data about students' clicks on the course website along with their engagement with Open Educational Resources (OER) found on BoxSand.org. We were also able to get click-stream data from the online homework system, Masteringphysics (MP) (Pearson MyLab & Mastering, 2019). These data are coupled to the course gradebook. Using statistical methods, we try to find interesting patterns in student's use of resources and how that use correlates with learning.

I was selected as an Ecampus Research Fellow with funding from the OER Unit at OSU Ecampus. I entered the fellowship with one academic year of data from a previously IRB approved research project. These data included students' usage of online resources and class performance. Prior to the fellowship, a preliminary exploratory analysis had been performed, but the project still needed more statistical rigor and modeling before publication. Therefore, the goal of my fellowship was to conduct a complete data analysis and produce a peer-reviewed manuscript. The motivations for this research and a review of our results will be discussed in turn below.

Background and Motivation

When I teach the Introductory Algebra-based Physics, my overarching student learning goal is the development of critical thinking skills to solve problems. Physics Education Research (PER) has shown that students learn problem solving and critical thinking best when they are engaged with peers while being guided by experts (Brame, 2013; Berrett, 2012; Crouch & Mazur, 2001; Lasry, Mazur, & Watkins, 2008; Rao & DiCarlo, 2000). To create an environment which best supports this mode of learning, I flipped my Physics class in 2014. Students now engage in pre-lecture videos, reading, and homework outside of class. During class students work individually and in small groups on a set of scaffolded questions that apply the fundamentals of physics. After class they have post-lecture practice and handwritten synthesis homework. By placing content delivery and tasks low on Bloom's Taxonomy (Armstrong, 2010), like root memorization, outside of class, more time can be dedicated to application and analysis while the instructional staff are present. When initially flipping my classroom, I was most concerned that students would not watch the pre-lecture videos to prepare for class. I made around 300 short videos that constituted a traditional lecture series and told the story of physics; if students didn't watch them, diving into problem solving in class could fail or be less productive than hoped. ***The curiosity to know what my students are doing in their outside-of-class study was the impetus of this work.***

To facilitate the flipped classroom structure, I built an Open Education Resource website, BoxSand.org (Walsh, 2016). The site consists of hundreds of pages with thousands of links to the best content found around the web and resources built by students and faculty at OSU. Content includes: videos, text, simulations, practice problems, infographics, concept maps, tips & tricks, and more. It is organized by topic and students are guided to what I and my colleagues think are the best resources. BoxSand tracks student clicks and video usage making it possible to know whether they watched the pre-lecture videos or did the pre-lecture reading. Over the past three years, depending on the time of the year and the lecture placement relative to an exam, about half to two-thirds of my students watched the pre-lecture videos before class. Very few read the textbook.

While building the hybrid structure of this class, I realized something profound – so much research and effort in education has been focused on crafting the best in-class experience. Whether a more traditional instructor-centered or progressive student-centered lecture experience, emphasis has largely been about curating the best class time experience (e.g., see Sorensen et al., 2006; https://www.physport.org/methods/). While this is important, many student handbooks claim students should expect to study outside of class about 2 to 3 times the number of hours spent in class. I certainly remember learning the most while doing homework with friends. It seems to me that many instructors and researchers spend the majority of their time and effort crafting what students do during only a small fraction of their learning (i.e. they focus on in-class time). The larger proportion of time spent studying outside of class is less structured. Usually, students are pointed to pages in a textbook and a few problems in the back of the chapter. While I was building my reformed curriculum, I spent much of my time talking with students to figure out what study habits and engagement patterns helped them succeed. I learned many of my students hadn't developed much in the way of healthy study habits. This too resonated with my experience of physics being the first class I had to really study for when I was a student. It took me significant time to develop my study methods and habits.

Realizing the importance of out-of-class guidance, my curriculum development efforts shifted to concentrating on a holistic path through the progression of a learning module. Everything is scaffolded and compartmentalized into small digestible chunks, centered on each lecture. This sets up a routine that guides students on how to best to learn. Combining this granulated curriculum with the ability to track student click-stream creates an opportunity to study instructional design choices in a way never before possible. My hope is to put more evidence into evidence-based instructional design.

The type of research my group engages in falls under the broad field of Educational Data Mining (EDM) (Baker & Inventado, 2014). We take in data from many different sources and look for patterns. We use statistical models to evaluate curriculum content and methods. In our first pass through the data collected on BoxSand, I was able to see that students largely do watch my pre-lecture videos. That gave me confidence that my class would not collapse if I lean on that mode of content delivery. As with most new research, answering one question leads to even more new questions. The next step was to ask not only whether students were watching the pre-lecture videos, but did it matter to their learning. Was there any correlation between watching pre-lecture videos and performance on assessments like exams? This type of EDM is called Correlation Data Mining (CDM) (Baker & Inventado, 2014). Much of the work of this project falls under the purview of CDM. Our basic research questions were:

1. ***Does engagement with pre-lecture videos correlate with performance on exams and overall grades?***
2. ***Does engagement with pre- and post-lecture homework correlate with performance on exams and overall grades?***
3. ***In general, what engagement with OER correlates with other engagement as well as with performance on exams and overall grades?***

During the study an emergent question arose about what should we be doing with all of this learning analytics research? Much of the time the research is only shared with instructors. I posed the question:

4. ***Does sharing learning analytics data with students increase motivation and encourage a more healthy engagement with the course?***

Results

PERC Proceedings Paper

I collaborated with Michael Dumelle and Katy Williams on our paper "*Tracking student engagement with open educational resources (OER) and online homework*" (Walsh, Dumelle, & Williams, 2019). It was accepted into the 2019 Physics Education Research Conference Proceedings

(PERC), a top PER journal. In the next few sections, I discuss four of the exploratory results and one of the model-based results presented in the paper.

Exploratory Analysis

I've chosen four exploratory results from our PERC paper to discuss here. The methods and results are explained in further detail in the PERC paper. I've also included additional analysis not found in the PERC paper.

Pearson Correlation. The first result from our paper was from an analysis of the Pearson correlation of a number of variables in our data. We wanted to know what engagement with OER resources and Mastering Physics website (MP) correlated with each other and grades in the class. This begins to address all three of our basic research questions by assessing whether a statistically significant correlation between variables exists. That helps us narrow in on what correlations warrant deeper analysis. We broke BoxSand click-stream data up into interactions with over 20 different types of content. Examples that show up in our final publication include watching videos, downloading solutions to homework and exams, engaging with simulations, reading the textbook, visiting the Overview or Tips and Tricks for a topic, and accessing the Calendar or Syllabus. We broke MP data into attempting pre/post-lecture problems and correctly answering those same problems. The only assessment metric we used was the overall course grade.

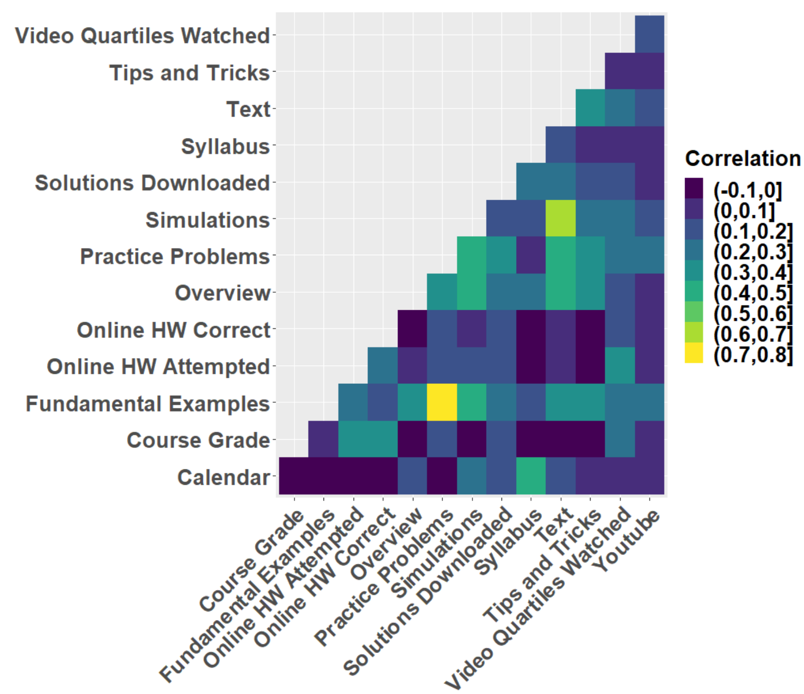

Figure 1. Pearson correlation between several BoxSand and online homework click-stream data sources and course grades. Each square represents the correlation between the corresponding variables on the x and y axes for 2017-2018.

As shown in Figure 1, we found a number of novel but small correlations among the use of different forms of BoxSand content. An example is that students who tended to engage in simulations were also more likely to read the textbook. We decided not to study most of these correlations further because of either their lack of statistical significance or their small effect size. For more details about our effect size threshold, see our PERC paper (Walsh et al., 2019). In the case of simulations and textbook reading, while the students that did one tended to do the other, the total number of students engaging in those resources were relatively low. With the MP data we saw that students who tended to watch pre-lecture videos were also more like to attempt the homework.

The most interesting data were how student engagement with these various resources and practice correlated with their performance in the class as measured by grades. Most of what we found would not be surprising to seasoned instructors, except for possibly that video usage positively correlated with grades, and the effect size of just attempting homework is relatively large. See the model-based analysis section below for more information on how large the effect sizes were and how they were determined. In contrast to our findings, previous work had shown that watching videos was not a predictor for student performance (Lin et al., 2017; Solli et al., 2018). We found there was a correlation between watching videos and grades. The story must be more complex**.** This did not surprise us as the pre-lecture videos are the main content delivery of the course and were created by us specifically for our class. Other studies have looked largely at video usage like YouTube, which were not created specifically for that course and often contain a narrative disjointed from the instructor. We confirmed this interpretation by looking at students who watched videos other than the custom pre-lecture videos (i.e. videos that were not created by us for our course). This engagement did not show any correlation with grades. We also found a positive correlation between overall grades and both attempting and achieving correct answers to MP questions. This was also not surprising to me, as I had learned physics by practicing problems. We need further research to determine if this is a universally beneficial study technique. When I transitioned from having pre/post lecture homework as suggested practice to requiring this homework as a part of student's grade, I found students actually did the work and the class averages increased noticeably.

Cramming. Inspired by the correlation of pre-lecture video watching and grades, we decided to dive deeper into general engagement with the BoxSand site and the videos. We looked at how usage evolved through time. The first noticeable result was discovered almost immediately.

Figure 2 shows the number of student sessions on the OER site BoxSand.org per day for an entire term (Google, 2017). This chart shows that activity increases during the middle of the week and spikes around exams. There is much less activity on the weekends and holidays.

We also found that the relationship between watching pre-lecture videos and grades depended highly on the week of the term. During off-exam weeks, watching more videos tended to correlate positively with higher grades on exams, while watching more videos during exam weeks had a negative correlation. This expands our understanding of the research question about how watching videos is correlated with grades by exploring the effect on a finer time scale than just a whole term and an overall grade.

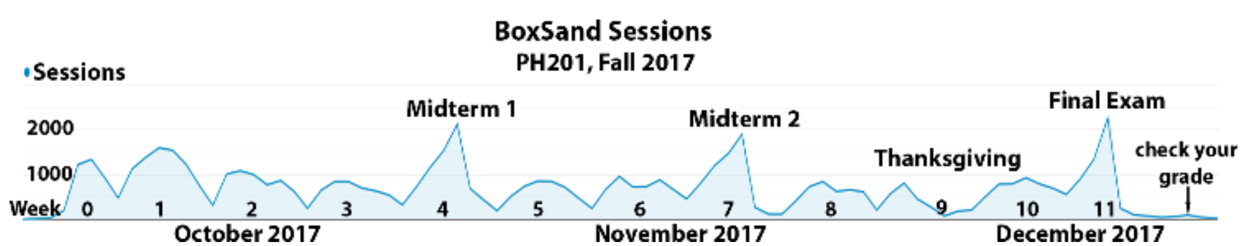

Figure 2. BoxSand sessions vs. time [13 time points] for fall 2017

We also found that the relationship between watching pre-lecture videos and grades depended highly on the week of the term. During off-exam weeks, watching more videos tended to correlate positively with higher grades on exams, while watching more videos during exam weeks had a negative correlation. This expands our understanding of the research question about how watching videos is correlated with grades by exploring the effect on a finer time scale than just a whole term and an overall grade.

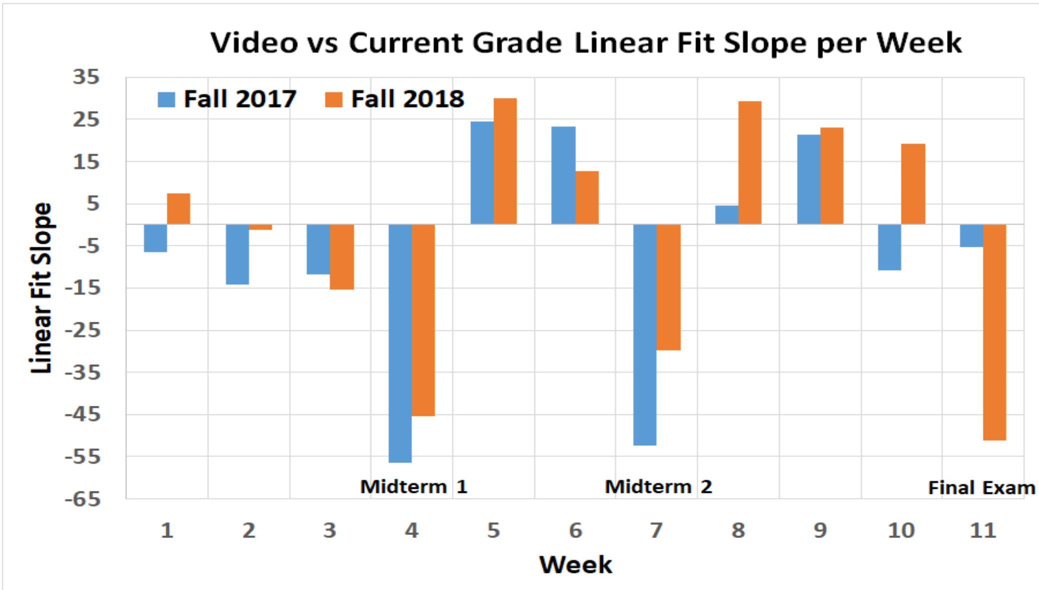

Figure 3. Linear regression slope for number of quartiles of BoxSand videos watched vs. current grade within each week for fall 2017 and fall 2018

Figure 3 shows the slope of a linear regression fit for each week of fall term. Positive values mean there is a positive correlation with watching videos and current grade. Negative values, as seen during exam weeks, show a negative correlation, which suggests that increased video use is related to lower grades. Note the plot may be misinterpreted as showing an overall negative correlation throughout the term but that is not the case, as the data in the plot is not scaled by overall use. While there is more activity during exam weeks, there are over three times as many off-exam weeks and the aggregate usage is skewed by the amount of data in those off-exam weeks. Further analysis on who is watching videos when, finds that the higher achieving students, who account for a larger portion of the video usage, watch videos more continuously throughout the term. Many lower achieving students, who don't make up the bulk video usage, tend to only watch videos on the days leading up to an exam.

Our interpretation of the data is that "cramming" for exams is less productive than steady engagement with course materials throughout the term. We believe watching videos or reading about a topic should be the first ~10% of a student's learning cycle while the rest should be application, practice, and synthesis**.** Leading up to an exam, students should be well past familiarization and introduction to the content. They should be practicing problems. This is largely confirmed when we look at student engagement with homework and how that correlates with exam scores each week. This pattern of the correlations flipping negative during exam weeks does not occur with homework. We interpret this pattern to suggest that practicing problems is a healthy way to prepare for exams.

Masteringphysics Engagement. To further explore the value of pre/post-lecture homework we looked at a distribution of student engagement with the Masteringphysics website (Pearson MyLab & Mastering, 2019) and course grades (see Figure 4). We found a clear trend where the students who are doing better in the course, shown further back in the three-dimensional plane in Figure 4, tend to complete a larger percentage of the pre and post-lecture homework overall, as shown moving to the right on the horizontal axis. The vertical axis shows the percentage of each student in each category. In the figure distribution, back and to the right represents the percentage of students who completed more homework and performed better in the class. Front and towards the left, represents the percentage of students who performed worse in the class and completed less homework.

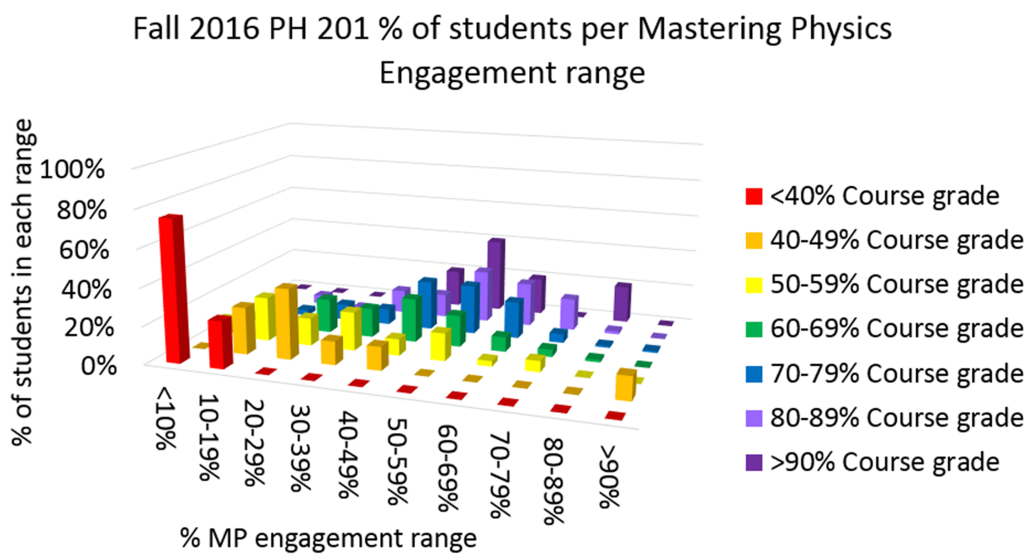

Figure 4. Percentage of each student in an online homework completion (engagement) range vs. course grade for fall 2016

It's important to note that full credit for MP assignments was set at a 67% correct rate, hence few students completed more than that. This figure also doesn't account for the number of students in each group. These results justify a more rigorous statistical analysis described in the model-based approach section below.

Grade Mobility. The last question our paper addressed with exploratory statistics was: *if we can detect a change in study behavior, can we detect a corresponding change in exam score*? To examine this, we looked at how much a student changed their video engagement from one exam period to the next. We then correlated that with the change in their exam performance. This directly addresses and expands our understanding of our research question about how video usage correlates with performance in the class.

Figure 5 shows the change in exam performance on the vertical axis. The value is calculated by taking the ratio of an exam score to the class average and then taking the difference between the values for two subsequent exams. For example, if a student received 60% on an exam that had a class average of 70%, then in the subsequent exam received a 60% but the class average increased to 80%, they would have decreased performance relative the class average. The change in exam performance would equal (60/80 – 60/70), or -0.107. This quantifies the change in how far from the average the students move. It is a figure of merit – a way to quantify the effect that may only have value when compared similar calculations. It attempts to account for exams with different class averages. The horizontal axis shows the change in the percent of the videos watched. A positive value of 1 means that whatever percentage of videos they watched in a previous time period, they watched twice that percentage during the second.

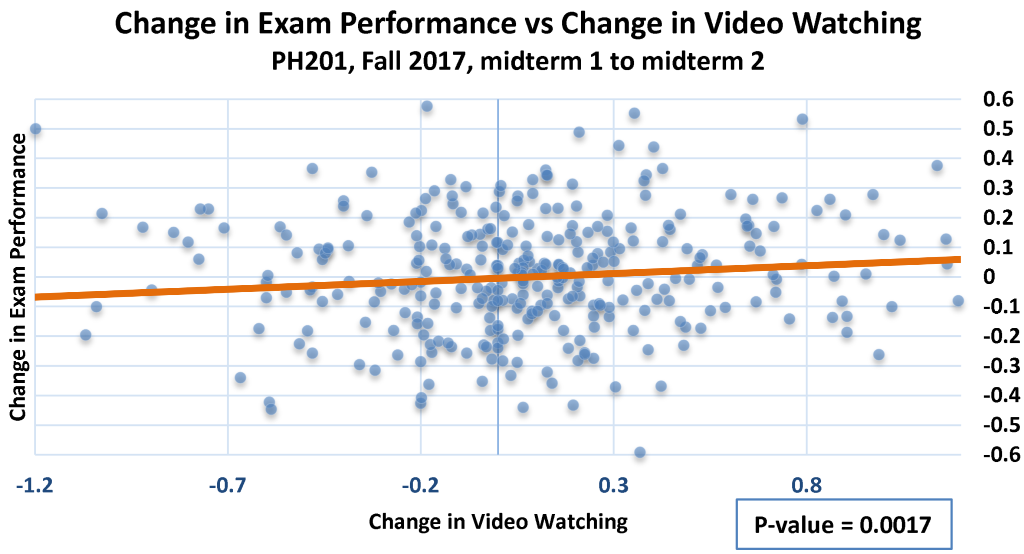

Figure 5. Change in exam percentage normalized to class average vs. change in video engagement percentage

We performed a linear regression and found students who increased their engagement with videos tended to also see increases in exam grades. While a positive correlation, the effect size was not considerably large – a student that doubles the percent videos they watch on average may see exam increases on average of less than 5% relative to the class average. We suspect this could be due to the limited nature of measuring student study behavior via one metric, video watching. Future study could include a more holistic measure of behavioral changes and possible show more significant correlations. We also found the effect size decreased over the year, but the correlation still remained positive. We suspect that was largely due to less variation in how students engage with the class. They tend to find a routine by the end of the first quarter and stick with it. It's also important to note that the act of normalizing the exam performance to the average is a non-linear process. Consequently, trying to interpret exact effect sizes from a linear regression on non-linear data is problematic. These results suggest more rigorous statistical analysis should be done on the effects of changing your study behavior.

Model-based Analysis

Linear Mixed Model. The exploratory analysis helped confirm the validity of our hypothesis about which variables correlated strongest with overall course grade. To test this hypothesis, we created a linear mixed model to quantify the effect size for these chosen variables. This addresses our research questions about how video usage and homework correlates to grades and allows the effect size to be quantified. Refer to our PERC paper for much of the details of this modeling work (Walsh et al., 2019). Here I will present a short summary.

We were pleased to find the Pearson test confirmed the variables of video watching, attempting homework, and answering homework correctly did correlate significantly positively with grades. Since one has to attempt homework to have a chance of getting it correct, we tested our model to make sure it didn't suffer from multicollinearity effects. Despite the feeling that getting a homework answer correct must be collinear with whether it was attempted, we found that not to be the case. Just because a student gets all the questions they answer right doesn't mean they attempted them all. One is not a predictor for the other. With this cleared, we performed a linear mixed model statistical regression. This allowed us to know how much each predictor variable contributed to the final grade in the class. To guarantee the predictor variable didn't directly contribute to the outcome variable, the points associated with pre and post-lecture homework was removed from the model. Our analysis then found:

1. Students who watched all of the pre-lecture videos had ~2.5% higher grade on average than those who watched none.
2. Students who attempted all pre/post-lecture homework had ~9% higher grade on average than those who attempted none.
3. Students who answered all their pre/post-lecture homework questions correctly had an additional ~10% higher grade on average than those who answered none of them correct.

We find these results intriguing because they confirm a long-held belief that homework in physics is highly valuable. If a student attempts all homework and gets half of them right, they tend to do ~14% better in the class than if they attempted none. In reality the pre/post-lecture homework contributes only 5% to your final grade. That would be nearly a 3-fold return on investment for that student. It also shows that while watching videos is correlated with performing better in the class, it has only a moderate effect. This is not surprising because you are more likely to learn physics by doing physics, than by just watching it.

Other Interesting Results

In this section I will discuss a number of results that did not make it into our PERC proceedings paper. Many of the results are preliminary with the intent of encouraging future work.

Early Engagement

When examining the correlations of some of the engagement predictor variables with the grade-based outcome, we saw student engagement behavior fluctuate early in the first academic term. The correlations and distributions changed each week, finally settling in on a consistent clear pattern by the middle of the term. This is concerning for two reasons: (1) students often need the motivation of the first exam to encourage them to develop healthy study behaviors, and (2) our current work involves predictive modeling to find struggling students early in the first term so that we can provide individualized support. We need students to see the value of engagement as early as possible. To see evidence of this effect, and why more research is necessary, we looked at the distribution of post-lecture homework engagement for a given exam period and grade on the 1st midterm. We compare that to the 2nd midterm, and all subsequent exams throughout the year. We present here the data about MP engagement and exam grade over time, but we saw similar patterns in all OER engagement. Therefore, these results can address the broader research question about how in general OER engagement is correlated with grades and how that correlation evolves through time, especially early in the course sequence.

Figure 6. Masteringphysics submissions versus exam grade for Fall 2016 grouped by course grade in 10% increments. The large circles represent the group average on that exam. The size of the circle represents the number of students in that group.

Figure 6 shows the MP question submission percentage and exam grade percentage for all three exams. The students are grouped by overall course grade in 10% increments, represented by the color of each data point. The average on an exam for a given group is represented by the large circles and the size of the circles represents the number of students in that cohort. The data show the distribution of the homework and exam grades changes early in the first term. The overall pattern varied, even on a weekly basis, until about halfway through the first term. At that point it became a steady state and remained relatively unchanged for the rest of the year. An example of this settling into a routine behavior can be seen by comparing the students who received a course grade below a

C, the red circle, which equates to less than a 50% in our class. At the 1st midterm, that cohort received a higher exam average than the next grade range up, the students who received between 50 – 60%, represented by the orange circle. By the 2nd midterm though their exam averages line up with the correct order of their course grades. This behavior isn't limited to just observing changes in exam averages. All cohorts started out completing more of the MP questions but decreased over time for the first 5 weeks until reaching a relatively steady state. The overall effect of the distribution of the circles is best shown when animated on a weekly basis. In general, the distributions of MP submissions decreased, the ordering of the averages aligned with a roughly positive linear slope, and the distribution on the midterm axis spread out. It did this for the first half of the first term until roughly reaching the bottom graph in the figure above. At that point it remained relatively steady for the rest of the year.

This early in the year fluctuation is seen when replacing MP engagement with most of the other variables we studied, including video usage, downloading solutions, accessing the syllabus, and engaging in OER on BoxSand. In some cases, the data allows finer time divisions and on a daily basis you can watch the fluctuations damp out and a trend that will last the rest of the year. While they all level out during the first term, some of the types of engagement get into steady state earlier than others. We think differentiating the time to steady state for various engagement with resources would be an interesting topic future research.

P-value Anomaly

We also noticed strange peculiarities about the statistical significance of some our data when sliced on a per week basis. For example, when you look at the correlation of video watching vs. grades on a per week basis (see Figure 7), you see during the first term on weeks 4 and 7 of a 11-week quarter the p-value is very low, indicating strongly significant correlations. These are midterm weeks, and recall from the earlier analysis on cramming, this is when watching more videos tended to be correlated with lower grades. The off-exam week p-values were not nearly as strong.

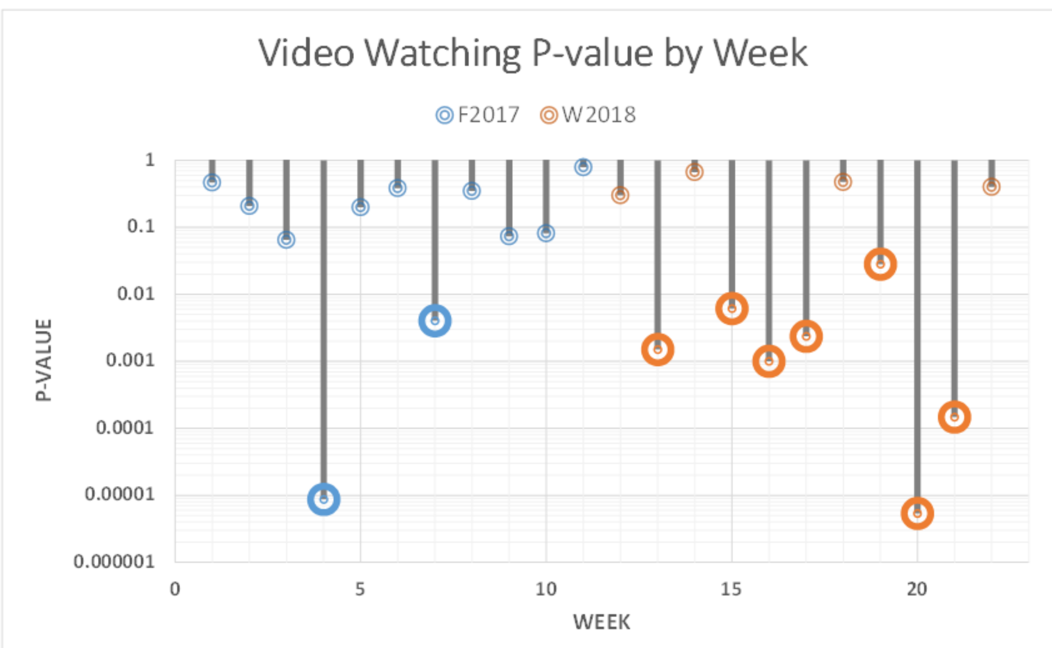

Figure 7. P-value by week for linear regression of video watching vs. course grade for fall (PH201) and winter (PH202) terms

In the second term, this behavior switches entirely and during the exam weeks, 3, 7, and 11, the p-values are not nearly as significant as off-exam weeks. This could be due to losing the weakest students between the two terms, but this is a relatively low number of students. What this illustrates is the need for further study and caution when trying to slice the data into finer time periods. There has to be a balance between wanting to see a current state of the system and needing large enough data points for statistical significance.

Other Correlations

We were a little disappointed that much of the content offered on BoxSand didn't appear to be correlated with higher grades. We thought that students reading the OER text or referencing some of the helpful sections like the tips & tricks or the topic specific problem-solving guide must be useful to their study. When looking broadly neither of these engagements showed significant correlation with grades. However, we also noticed that very few students actually visited any of these resources. This is a continual lesson that if it's not assigned for points, and is considered supplemental by the students, they will not engage. To test if the resources were helpful to those who did engage, we removed everyone who hadn't used the resource from the data set. Then we saw a more promising pattern.

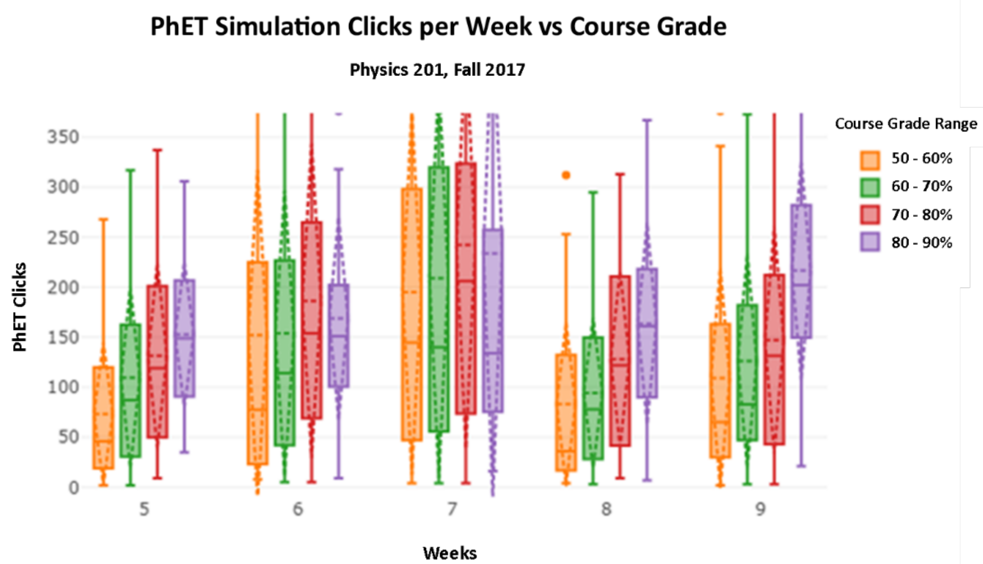

Figure 8. PhET Interactive Simulation clicks per week, for students who clicked at least once, grouped by course grade with 10% intervals

Figure 8 shows that of the students that visited the PhET Interactive Physics Simulations (Perkins et al., 2006) the more they engaged with the resource, the higher their course grades tended to be. The plot shows the median number of clicks represented as the horizontal line near the middle of each box, and the first and second quadrants represented by the ends of the box and the error bars, respectively. The center of the triangle represents the mean, while the triangle ends represent the standard deviation. Students are grouped by course grade range in 10% intervals. It is important to note the upward trend shown during weeks 5, 6, 8, and 9. Increased clicking on the simulations tended to be correlated with higher grades. This correlation is weakened during week 7, which was an exam week. This is consistent with the results from the cramming section reported earlier, where the slope of video watching and grades changed drastically during exam weeks.

These plots looked essentially the same when analyzing other resources that didn't correlate with grades in the aggregate but did when limited to the subset of students that used the resource. Perhaps we are just seeing that the students willing to click around to a more diverse set of resources are the more motivated students who perform better in the class. I don't think this necessarily tells us the value of specific content. Perhaps we are seeing that these resources are useful in learning. Either way, more research is necessary. However, these results help to address our research question about general OER engagement and how it correlates with grades by eliminating missing data.

Fine-grained Assignment Impact Analysis

One question I have always had is how much of an intervention is necessary to see a statistical difference in the outcome? *Can we quantify the effect of individual activities and engagement with specific OER as they are added to the curriculum?* This could be used to systematically test the efficacy of individual assignments or even specific problems. If engagement doesn't correlate with learning, then it is just busy work and should be removed from the curriculum. To study this, we performed a pilot test during spring term in 2017. We created two activities centered on different simulations. One of them was part of the students' challenge homework, making it more compulsory. The other was a suggested activity worth no points. We then planned an exam question to assess whether the activities were related to their grades. As you might have expected, very few students engaged in the activity that was only suggested, so we didn't continue testing the students further on that skill. We did continue with the analysis for the simulation that was part of a homework assignment. This addresses the research question about what general OER engagement correlates with grades by studying individual pieces of content and seeing if their influence can be measured.

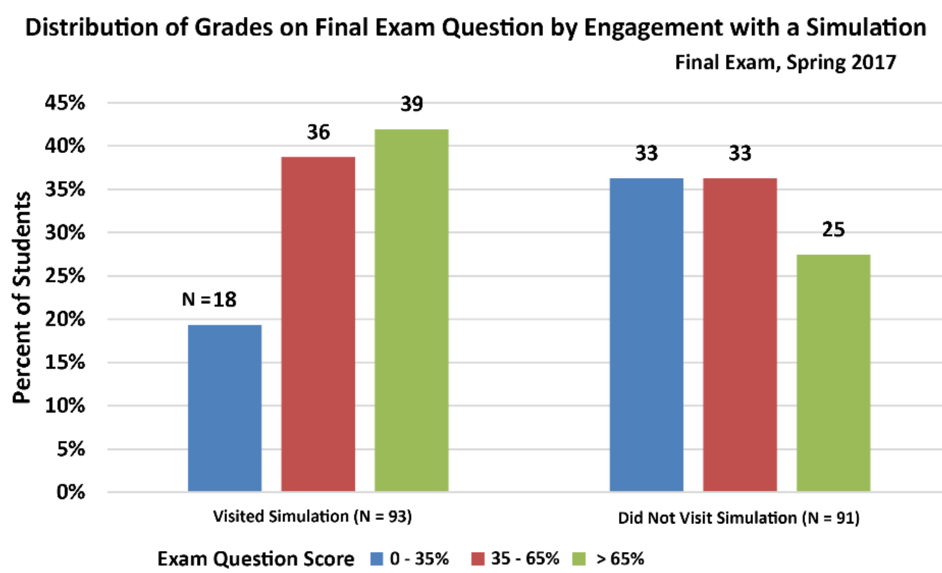

Figure 9. Percent of students that received between 0 – 35%, 35 – 65%, and 65 – 100% on final exam question with corresponding homework question that involved interacting with a simulation, grouped by whether they visited the simulation or not

We divided the students up by whether they got 0 – 35%, 35 – 65%, and 65 – 100% on the exam problem. Figure 9 shows the students that did visit the simulation and do the homework problem associated with this test, tended to do much better on the corresponding exam problem. The largest percentage of students in this group received more than two thirds of the points and the smallest percentage received less than a third. For the students that did not visit the simulation there was no clear pattern, and the smallest percentage of students received over two thirds of the points on the problem. Like much of this research, we are likely looking at students that are generally higher achievers and thus do all their work versus those that do not. That said, it does help shed light on how small of an exercise can have a measurable outcome. You can have one homework problem out of the nearly 20 a student might see in a given exam period and find a noticeable outcome in exam performance from their engagement with it. You don't have to poke the beast very hard to get a reaction.

What to do with this data?

Motivating Students

I think one of the most important questions is what to do with all this information? An interaction with a student helped convince me that one thing we can use this data for is better communication about good study habits**.** One morning I questioned: *How does downloading solutions to homework and exams correlate with grades?* I rushed to campus and made a simple plot of the average number of downloaded solutions for each of the grade ranges (see Figure 10).

As I finished, office hours began, and the first person who arrived was a very dedicated student. While this student mostly received C's in the class, I would by all means consider them successful; their knowledge gains on both directly and indirectly assessed learning outcomes were quite high. They just came into the physics series relatively ill-prepared due to circumstances outside of their control.

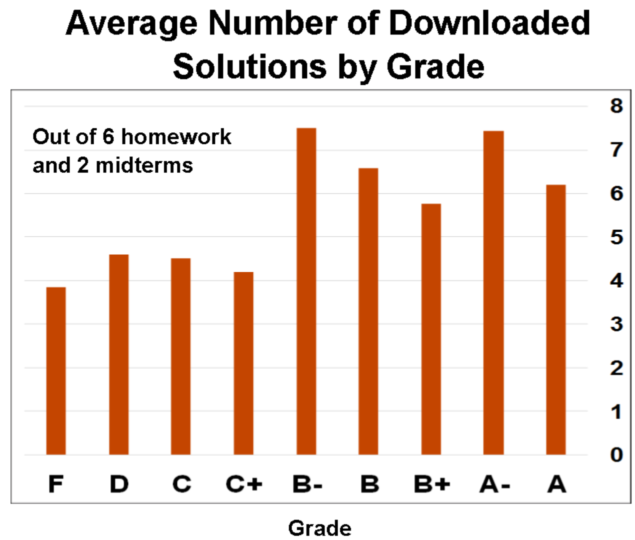

Figure 10. Average number of downloaded homework and exam solutions grouped by course grade for PH201, Fall 2016

Before helping them on physics I wanted to show them this plot; I had a relatively good rapport with this student. I said, "Look at that cliff between the A's and B's and the C's and D's". Clearly the higher achieving students were downloading more of the solutions than the lower achieving students. This student then immediately pointed to the C range and said, "I'm here, does that mean I should be downloading and reviewing the solutions more often"? I went through about 6 emotions in a two second period. How many times had I expressed to the class the importance of this metacognitive stage in learning? But I ended feeling proud. I am teaching science majors and part of the larger goal is to teach them to use data and logic to derive their own conclusions. They won't always have a professor nearby to guide them. They need to move past defaulting to authority and take ownership in their own analytical skills to make decisions. I realized at that moment that we need to get these data about class performance in relation to resource engagement into the students' hands.

I now spend the first day of class not talking about the syllabus or the course, which is all available in videos and text online. Rather I give what would amount to as a motivational presentation called "*Why we are here*". I finish the presentation by discussing Physics Education Research and the lessons it has brought us about students learning problem solving and critical thinking best by talking with peers and guided by experts. This all is couched in a way to motivate acceptance into the flipped classroom structure and help them understand why I'm asking them to work in ways most have never experienced. I finish with data from our class, some of which has been shown here. My hope is they will use this information to make a conscious change to how they approach learning, studying, and this class. I think this is time very well spent. Students come out of the talk excited to be doing physics and confident in the reasons for a new classroom model. Years later, many refer back to that day and the impact it had on them.

Evidence also supports the *Why we are Here* presentation provides early motivation to students. Figure 11 shows the total number of BoxSand sessions for the first 3 days of Fall term. The number of sessions is scaled to the number of students to account for different enrollment numbers. When looking at activity on the BoxSand site for those first few days, early engagement increased between fall of 2016 and fall of 2017 by nearly 30%. The only difference to the start of the term was the motivational presentation with data specifically from our class.

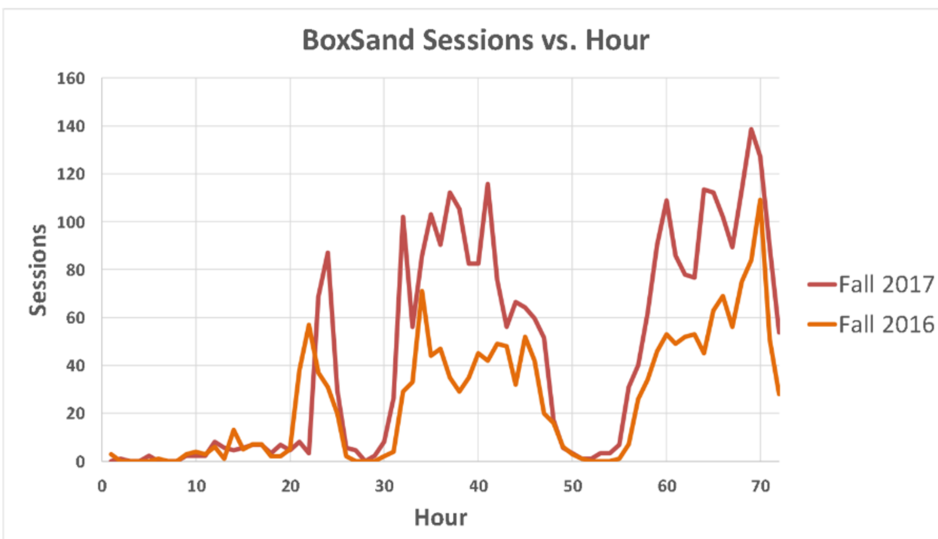

Figure 11. Total number of new BoxSand sessions per hour, scaled to the number of students, for the first 3 days of Fall term

The increased engagement during fall of 2017 continued to be larger than the previous year but the difference decreased over time. If motivating students with data gets them started earlier and overall results in more engagement, that is a good thing. Almost all of our data show that increased effort results in increased grades. These results helped answer the last emergent research question about whether sharing learning analytics data with students would encourage a more healthy engagement with the course by at least showing it increased engagement. Future work needs to explore what was the nature and quality of that extra time engaged with the course resources.

The success of sharing learning analytic data with students doesn't surprise me. We live in the age of the quantified self. Fitbits track and report statistics on our health. Most banks provide incredibly detailed analytics on how you spend your money. Increasingly more students expect these kinds of analysis. I've had a student tell me there were shocked and a bit offended that we are just now starting this kind of tracking. They questioned how anyone made curricular decisions without these kind of data in hand. They were frustrated that companies will track their every movement just to find ways to sell more widgets, but we aren't using this technology to support more positive goals like learning. I won't go so far as to say that you can't make good decisions without analytics, but I do think it should be a part of evidence-based instructional practices going forward.

Perspective

The last thing I will say is a bit of warning about what this research is and what it is not. Educational science, which is inherently a form of social science, is notoriously squishy. Conclusions are rarely definitive. The array of confounding variables is innumerable. I don't claim that the type of quantification shown in this paper, like the results from the linear mixed model, are a complete picture. This does not mean quantitative education research does not provide good evidence, it simply means you must understand the limits of its conclusions. One major criticism I have of this work is that it relies on using grades as the only assessment of student learning. I've tasked my research group with thinking more holistically about measuring learning outcomes. Our immediate efforts are to divide these results based on demographics and see what conclusions appear more generalizable than others. This will inevitably generate new questions, which is possibly the greatest value in this work. Quantified PER is really good at finding new questions. While it provides evidence to support curriculum changes it doesn't necessarily peer into the underlying mechanisms driving the observed effects. Comparing it to medicine, we don't always know why a treatment provides a more positive outcome, but we still employ the preferred method while we learn the intricate details. Coupling this type of quantitative education research with more traditional qualitative approaches can create a healthy synergy. Quantitative researchers find interesting questions that are then handed to deep cognitive scientists who are better equipped at answering them, all the while continuing to make strides in more positive learning outcomes.

Acknowledgements

I want to thank undergraduate researchers Ikaika Mckeague-Mcfadden, James May, and Jake Bigelow for their contributions to this work.

References

Armstrong, P. (2010). *Bloom's taxonomy.* Vanderbilt University Center for Teaching. https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/

American Association of Physics Teachers. (n.d.) *PhysPort: Teaching methods and materials.* https://www.physport.org/methods/

Baker, R. S. & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp.61–75). Springer.

Berrett, D. (2012, February 19). How 'flipping' the classroom can improve the traditional lecture. *Chronicle of Higher Education.* https://www.chronicle.com/article/how-flipping-the-classroom-can-improve-the-traditional-lecture/

Brame, C. (2013). *Flipping the classroom.* Vanderbilt University Center for Teaching. http://cft.vanderbilt.edu/guides-sub-pages/flipping-the-classroom/

Crouch, C. H. & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, *69*, 970–977.

Google (2017). Google Analytics. https://analytics.google.com/analytics

Lasry, N., Mazur, E. & Watkins, J. (2008). Peer instruction: From Harvard to the two-year college. *American Journal of Physics*, *76*, 1066–1069.

Lin, S. Y., Aiken, J. M., Seaton, D. T., Douglas, S. S., Greco, E. F., Thoms, B. D., & Schatz, M. F. (2017). Exploring physics students' engagement with online instructional videos in an introductory mechanics course. *Physical Review Physics Education Research*, *13*(2), 020138.

Pearson MyLab & Mastering (2019). *Mastering physics*. http://pearsonmylabandmastering.com

Perkins, K., Adams, W., Dubson, M., Finkelstein, N., Reid, S., Wieman, C., & LeMaster, R. (2006). PhET: Interactive simulations for teaching and learning physics. *The Physics Teacher*, *44*(1), 18-23.

Rao, S. B. P. & DiCarlo, S. E. (2000). Peer instruction improves performance on quizzes. *Advances in Physiology Education*, *24*, 51–55.

Solli, R., Aiken, J. M., Henderson, R. & Caballero, M. D. (2018, August 1-2). *Examining the relationship between student performance and video interactions.* Physics Education Research Conference, Washington DC. doi:10.1119/perc.2018.pr.Solli

Sorensen, C. M., Churukian, A. D., Maleki, S., & Zollman, D. A. (2006). The New Studio format for instruction of introductory physics. *American Journal of Physics*, *74*, 1077-1082.

Walsh, K. (2016). *Project BoxSand.* http://boxsand.org

Walsh, K. C., Dumelle, M. & Williams, K. (2019, July 24-25). *Tracking student engagement with open educational resources (OER) and online homework.* Physics Education Research Conference, Provo, UT. https://www.compadre.org/per/items/detail.cfm?ID=15299

About the Research Unit at Oregon State Ecampus

Vision
The Ecampus Research Unit strives to be leaders in the field of online higher education research through contributing new knowledge to the field, advancing research literacy, building researcher communities and guiding national conversations around actionable research in online teaching and learning.

Mission
The Ecampus Research Unit responds to and forecasts the needs and challenges of the online education field through conducting original research; fostering strategic collaborations; and creating evidence-based resources and tools that contribute to effective online teaching, learning and program administration.

Contact
Mary Ellen Dello Stritto, Ph.D.
Director of Research
Oregon State Ecampus
541-737-4697
maryellen.dellostritto@oregonstate.edu

Creative Commons License

Suggested Citation
Walsh, K. C. (2020). Project BoxSand: Tracking student engagement with open educational resources and homework. *White Paper*. Corvallis, OR: Oregon State University Ecampus Research Unit.