Predicting on-campus student performance from video interactions

Robert Solli*, John M. Aiken*, Rachel Henderson†, Marcos D. Caballero*† §
*Center for Computing in Science Education & Department of Physics, University of Oslo, N-0316 Oslo, Norway
†Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824
§CREATE for STEM Institute, Michigan State University, East Lansing, Michigan 48824

MICHIGAN STATE UNIVERSITY

UiO : Department of Physics
University of Oslo

Research question

How does student interaction with tutorial videos relate to the performance on laboratory exercise presentations in an introductory mechanics course?

Background & Motivation

- Video lectures are implemented as a common tool in physics classrooms
- "In-video" drop out rates vary according to production quality, length of video[1,3,4]
- Previous research has shown that student-video interactions have predictive power on student performance, as measured by an "in-video" quiz[6]
- On the same cohort of students, Lin *et al.* explored viewing behaviour which showed a consistent fraction of students accessed laboratory videos[1]. This allowed us to use the same features for our model across all laboratory activities.

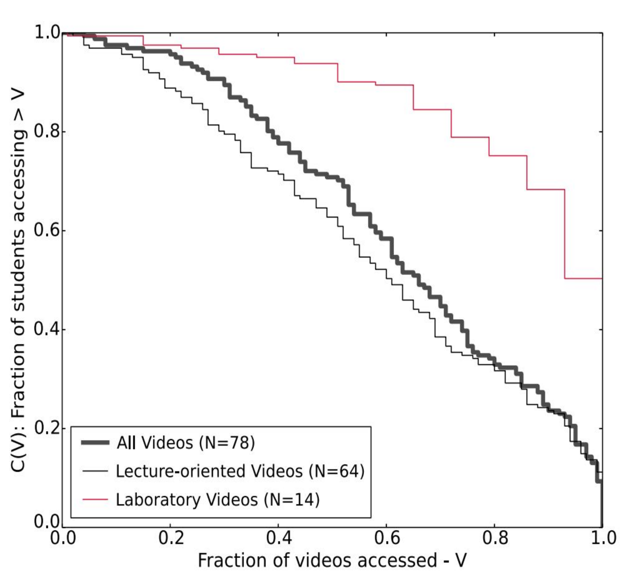

Figure 1: Student access of different types of videos during the course. Laboratory videos retained a high percentage of access relative to lecture videos.

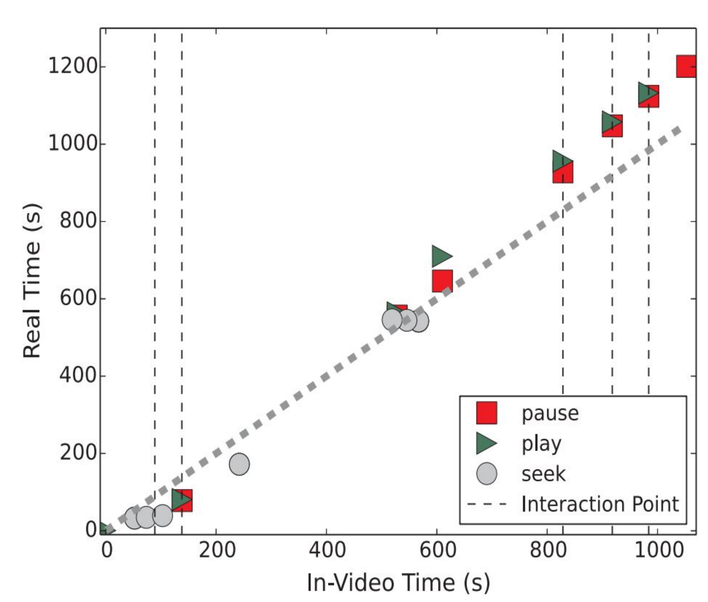

Figure 2: An individual student's behaviour with a video. The diagonal dashed line represents where real time = video time

The Course

- On-campus students in an introductory, calculus-based mechanics course in fall 2013
- Offered as a flipped classroom at Georgia Institute of Technology with N = 161 enrolled students
- Matter & Interactions curriculum[7]
- Content material throughout the course was supplied in the form of videos (78), 14, of which, were associated with the laboratory activities.

References:

[1] S.-Y. Lin et al., Phys. Rev. Phys. Educ. Res. 13, 020138 (2017).
[2] R. S. Olson et al., arXiv preprint arXiv:1708.05070 (2017).
[3] P. J. Guo et al., in Proceedings of the first ACM conference on Learning@ scale conference (ACM, 2014) pp.41–50.
[4] J. Kim et al., in Proceedings of the first ACM conference on Learning@ scale conference (ACM, 2014) pp. 31–40.
[5] D. T. Seaton et al., European MOOC Stakeholders Summit, Lausanne , 140 (2014).
[6] C. G. Brinton and M. Chiang, in 2015 IEEE Conference on Computer Communications (INFOCOM) (2015) pp. 2299–2307.
[7] R. W. Chabay and B. A. Sherwood, Matter and interactions (John Wiley & Sons, 2015).

Data collection

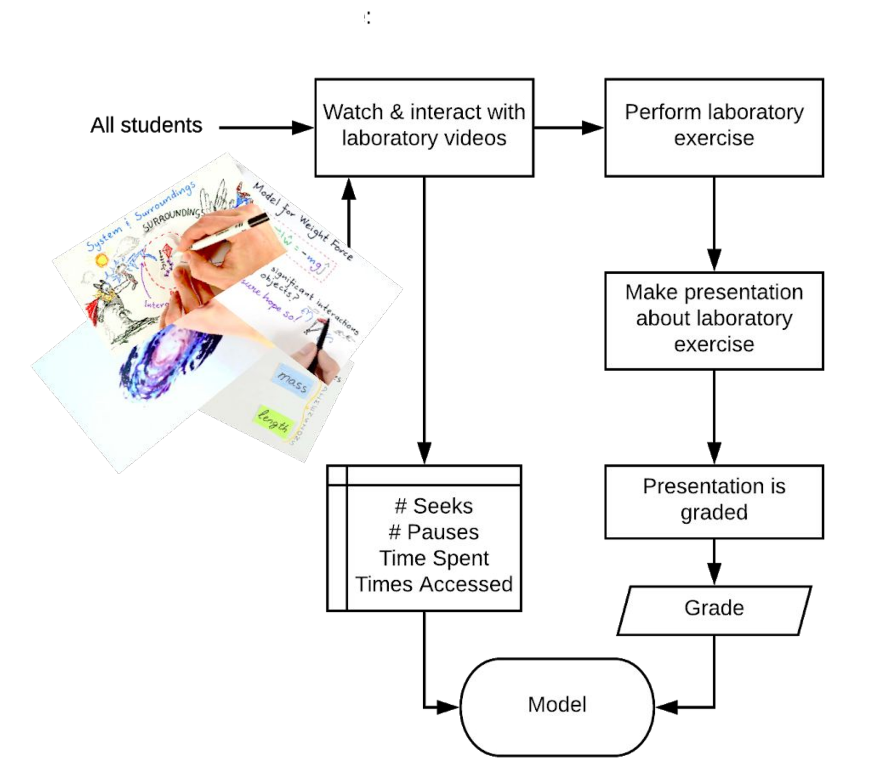

Figure 3: For each of the four laboratory exercises, the students watch laboratory videos and their interactions with these videos become the input features to our model (see Figure 2 for an example). See Table 1 for a description of the features. After watching the videos the students perform a laboratory exercise which is graded - and the grades are the outcome variable for our model.

Results

- After a linear regression model failed to predict student grades on laboratory activities (R2 = 0.056), we opted to cast the problem as one of classification.
- Aim: To classify students as performing above or below the population median for a given laboratory activity.
- Features from Table 1 served as input to a Logistic Regression (LR) model
- Both LR1 and subsequent LR2 performed similar to the trivial random-guess classifier

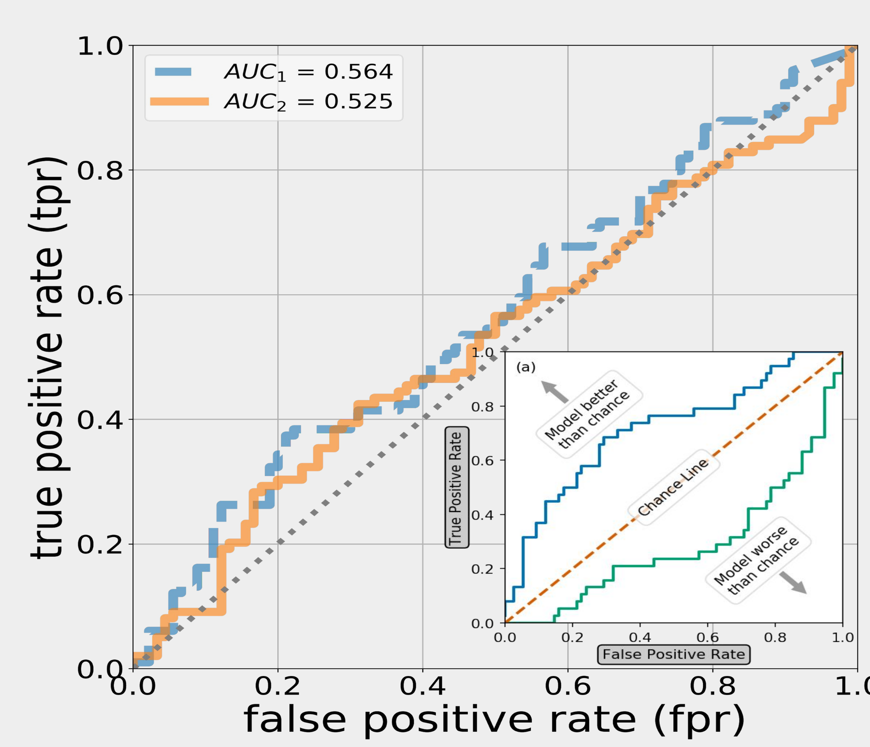

Figure 4: Two models were constructed: LR1 (D_n = 0.118, p = 0.335) and LR2 (D_n = 0.084, p = 0.772). The additional contextual information added to LR2 did not increase model performance.

Features

Feature name	Description
t_N *	Time each student spent watching a video, normalized by the video length, averaged - Z-scored
t_A *	Number of individual accesses to each video averaged Z-scored.
pauses *	Number of pauses with automatic pauses removed in the video averaged - Z-scored.
plays *	Number of plays averaged - Z-scored.
seeks *	Number of seeks averaged - Z-scored.
interaction time	Average time relative to the population mean of the student interaction with videos, averaged - Z-scored
lab	Which lab the feature vector corre-sponded to in the course sequence
$FMCE_{pre}$	Each students score on the FMCE test prior to taking the course.

Table 1: Measures of student interactions with tutorial videos as well as contextual information: proximity to deadline, course sequence, FMCE pretest. Features denoted with an asterisk were used in the first model, denoted LR1. Contextual information was added to LR1; this model is denoted as LR2.

Model Evaluation

- The model is *trained* and then *validated* on separate parts of the dataset (see Figure 6)*Validation* serves to estimate generalization error. We use the Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) to quantify model performance[5]

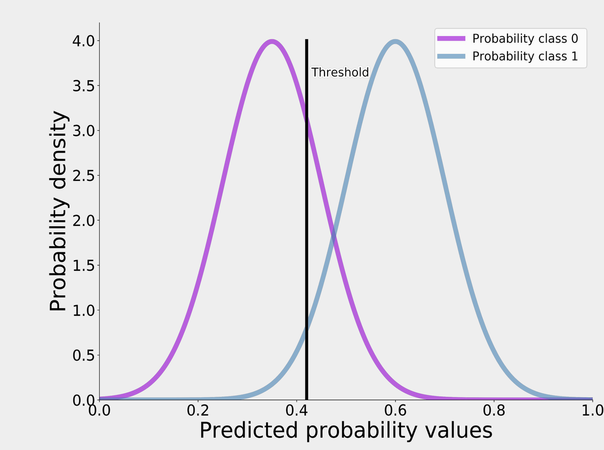

Figure 5: A classifier is good if the mean of the distributions of probabilities are far apart (with the deviations relatively similar). As the threshold slides over the values in [0, 1] points in the ROC curve are by the *Sensitivity* and *Specificity* for each threshold value

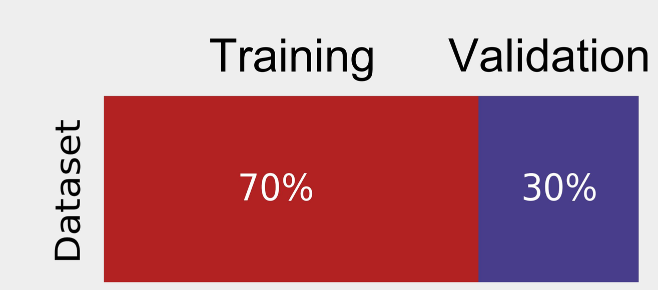

Figure 6: In Machine Learning, to avoid *overfitting* the data is randomly segmented in *Training* and *Validation* sets. Our split is 70/30, a common value [2]

Conclusion

There does not appear to be a connection between how students use laboratory tutorial videos and student performance on more complex laboratory activities

In summary our study suggests:

- Although the models have AUCs of > 0.7 have been shown[6] for predicting performance on end-of-video quizzes the complexity of the laboratory assessments might be too high
- Including contextual information, like FMCE pretest scores, relative interaction time and course sequence did not improve the model
- Videos in a flipped course might not serve the precise purpose that the designer intended and that more intentional and innovative designs might be needed to be explored.
- That a good model of student performance on complex activities, like the laboratory assessments in this course, likely starts from other forms of engagement in the course

Acknowledgments:

This study was funded by the Olav Thon Foundation and the Norwegian Agency for Quality Assurance in Education (NOKUT), which supports the Centre for Computing in Science Education.