IMPACT OF ONLINE RESOURCES IN TEACHING STATISTICS TO UNDERGRADUATES

Elinor M Jones, Simon Harden and Jian Yu
Department of Statistical Science, University College London, London, UK
elinor.jones@ucl.ac.uk

We investigate the impact of a virtual learning environment (VLE) on the results of 190 undergraduate students, mostly statistics or mathematics majors, enrolled on a course in applied statistics. The VLE provides access to a variety of tailor-made resources varying in degrees of interactivity, including videos of lectures, and is intended to supplement weekly face-to-face lectures. Student engagement with the VLE is automatically tracked. We investigate what VLE resources students use and when, and whether academic background impacts on how these are used. Finally, a model for academic achievement based on student interaction with the VLE is built.

INTRODUCTION

Teaching in Higher Education has gone through a digital revolution in the last twenty years, with students now expecting more than passive lectures. Part of this shift has been driven by the use of Virtual Learning Environments (VLEs): purpose built central repositories for course materials and other relevant resources, as well as a platform for interactive or multimedia resources such as videos, quizzes and discussion forums. Coupled with increasing availability of lecture recording technology allowing students to watch lectures online in their own time, this provides the ultimate flexibility in deciding how and when to learn.

There has been considerable interest in how students use these alternative study resources, and which they find most helpful. Since most VLEs provide some level of automatic tracking of student activity, such data has been used for course re-design (Baepler and Murdoch, 2010), as an early warning system to flag disengaged students (Macfadyen and Dawson, 2010), or to understand the relationship between achievement and engagement with VLEs (Demian and Morrice, 2012).

We consider the use of a VLE by students enrolled on an Undergraduate course in Statistics. The course has no compulsory attendance (actual attendance ranged from 50-75% in any given lecture, of which there were three hours per week) or coursework, barring a short in-class test toward the end of the teaching term and the final exam the following summer. Under such circumstances, understanding how and when students use resources is of interest, along with understanding the implications for overall achievement. This culminates in a linear model for student understanding of the material, measured using their final mark for the course, on derived measures of a student's activity on the VLE between the start of the course and the final exam.

THE DATA

The data for this paper come from students enrolled on a second year course in Linear Models and the Analysis of Variance at University College London, UK, for 10 weeks in the Autumn term of the academic year 2016-17. The class size was 190, with the majority of students enrolled on a degree in the Department of Statistical Science (123), while others were completing degrees in the Department of Mathematics (40) or in other sciences (27).

The VLE for this course is Moodle, a flexible and popular platform in Higher Education, and all students enrolled on the course were given access to the course Moodle page. The resources available therein include all materials used in class, in addition to a large number of optional or extra-curricular activities. The resources and activities on the Moodle page are grouped according to type and displayed neatly in separate, well-labelled sections, designed to make it easy and intuitive for students to navigate and find relevant information.

Resources available to students

The resources available include lecture slides and a full set of course notes which covers all course material. Exercises were released each week on Moodle, and though there was no obligation for students to complete them, they were encouraged to do so before they were discussed in a lecture the following week; solutions were made available following this session. Examples of past examination papers and class tests, with sketch solutions, were available, in addition to 'feedback' documents detailing the common mistakes made by the students who sat these assessments.

Additional resources included recordings of all lectures and two revision videos created using the Articulate Storyline software. The latter are animations where students are able to interact directly with the material on-screen, and try embedded quizzes with in-built feedback to check understanding. Students were also encouraged to learn the software R using bespoke worksheets created for the course. Finally, ideas for extra-curricular activities to encourage and motivate students to explore the subject area beyond the scope of the course were provided in the form of links to interesting and relevant webpages.

In addition to resources, the Moodle page hosted a forum and students were encouraged to post questions, anonymously or otherwise. Students have the capability to answer their peers' forum questions, though the majority of the questions are answered by the lecturer.

Capturing student interaction with the VLE

All activity on the course Moodle page is automatically tracked from the start of the course in October 2016, through to the end of the course in December 2016, and beyond until the final exam in May 2017. Moodle records the time, date and nature of each 'action', for example, accessing a particular resource, following a URL to a suggested site, or simply moving between sections of the Moodle page.

This provides a rich resource for tracking student engagement with the provided material. It allows us to ascertain *which* resources students are keen to use, and *when* resources are accessed by students. Much of the material available, however, is downloadable: notes, slides, exercises, solutions are all pdf files, and students are expected to download and print these to use during the term. Beyond the initial access point, students may not access these resources online again, relying on a downloaded or printed copy. Tracking usage of such resources over time is therefore not possible, though the timing of initial access to the resource is interesting in itself.

A similar issue occurs for other activities and resources such as viewing posts on the forum and URLs to external sites. For the former, all students are subscribed to the forum mailing list so forum posts can be read via email rather than visiting the site; posting to the forum is another matter – students must visit the course page in order to do this.

Other material, such as lecture recordings and Storyline videos could be downloaded by students but the file sizes are so large that only a handful of students choose to do this; the majority view these online and so activity is tracked. Each access point to Storyline videos is recorded, though interaction within this resource cannot be tracked. However, lecture recordings are stored on a separate system, Echo360, with a link to this resource available from the course Moodle page. Echo360 records student interaction with each of the recorded lectures, including the number of videos accessed, how many times a particular video was watched, and the average percentage of the video watched in a single sitting.

In order to study student interaction with the material, resources available on Moodle were considered as either: (i) *essential and downloadable material* (e.g. lecture slides, notes); (ii) *non-essential material which are downloadable or viewable elsewhere* (e.g. forum posts, URLs); and (iii) *non-essential and non-downloadable material* (e.g. lecture recordings, Storyline videos).

Since it is not possible to track the usage of the types of resources listed in (i) and (ii) above, we look at first access points to material under type (i), and interaction with resources in type (ii) where possible (e.g. first use of a URL, posting to the forum). Resources of type (iii) can be tracked throughout the teaching term and up until the final exam for the module.

RESULTS – ACTIVITY LEVELS ON MOODLE

All students accessed the Moodle page within two weeks of the start of the course, and many had accessed the Moodle page up to two weeks prior to the start. Figure 1 shows the number of students accessing the page from the start of the course through until the exam (blue line). During term time, spikes occur on the day of a lecture, the day before the class-test at the end of November, before rapidly dwindling for the rest of term. Access during the second term, when there is no formal teaching for this course, is sparse, though picks up at the start of the Easter holidays before the exam. The last spike occurs the day before the exam. No differences in access to Moodle was detected between students from Statistical Science, Mathematics or other sciences.

Type (i) resources

Lecture notes are supplied for the course as a 130 page pdf document. These contain all material covered in the course, though discussions with students reveal that not all read this document – many only dip in if specifically instructed to do so. The majority of students accessed the lecture notes for the first time sometime during the first five weeks of the course.

Lecture slides were released in sections, with six non-overlapping pdf documents between them containing all slides used in class. The first three batches were accessed in good time during the term – perhaps motivated by the in-class test at the end of November which tests understanding of the material contained therein. However, for the last three batches of slides, an increasing proportion did not access the material during the course. Notably, around half of all students did not access last set of slides until the weeks leading up to the exam. This pattern of behavior tallies with falling rate of lecture attendance through the term.

A total of six exercise sheets were released in total throughout the teaching term, and solutions to these were discussed in a tutorial approximately a week after their release. Surprisingly, only 13 of 190 students accessed all six exercise sheets in a timely manner before the relevant lecture. One hundred students accessed two or fewer exercise sheets in a timely manner.

Unsurprisingly, most students accessed past papers and relevant solutions, to both the in-class test and the final exam, in a timely manner.

Type (ii) resources

This includes extracurricular activities such as tutorials in learning to use R, URLs for various statistics pages, and use of the forum. A total of 122 students accessed the URLs at some point during the course, with the majority doing so either in the first few weeks of the course, or in the days leading up to the final exam, and a similar pattern for accessing R tutorials is apparent.

A total of 141 posts were made to the forum, 51 during term time (with 44 of these in the two weeks leading to the in-class test) and 90 after the course had finished (with 68 of these in the two weeks leading to the final exam). These are mostly questions, though some were clarifications or students responding to each other. Most forum posts were answered by the lecturer.

Type (iii) resources

Figure 1 also shows the number of students accessing the lecture recordings. Of the non-downloadable materials, this was by far the most popular resource (167 of 190 students accessed these recordings at least once), along with the two Storyline revision videos (169 of 190 students accessed at least one recording). The same pattern of usage is observed for lecture recordings as access to Moodle – this is unsurprising considering that the vast majority of students watch the lecture recordings on-line as opposed to downloading them. However, far more students access Moodle than those accessing the lecture recordings; other resources therein must also be of interest to students. This is understandable during term time as new lecture slides and exercise sheets are added to the Moodle page as the term progresses, but this is not the case in the few weeks prior to the exam.

While usage patterns of most resources don't differ hugely depending on home department, use of multimedia resources such as lecture recordings and Storylines revealed significant differences. Mathematics students were the least likely to watch the Storylines, with only 20% of students accessing *both* videos before the exam; this is in contrast to 45% of Statistical Science and 63% of other science students doing so. The pattern is reversed for lecture recordings – Mathematics students tend to watch more unique recordings, watch them more often, and view more of each recording in one sitting (Kruksal Wallis *p*-value approximately 0.05 for all three). Why a difference between Mathematics students and other disciplines exists is not known.

RESULTS – RELATIONSHIP BETWEEN ACTIVITY LEVELS AND ACHIEVEMENT

Achievement at the end of the course is computed using their mark from a written mid-term test (weighting of 10%) and a written summer exam (weighting of 90%). No significant difference between those from Statistical Science, Mathematics or other sciences were found in terms of overall achievement.

Data from lecture recording usage was available as a summary for each student, and was merged with the wide-format data created from the original Moodle data. To assess how overall mark relates to activity on Moodle, variables of interest were derived from these data. Though these derived variables are subjective, they were deemed of interested based on other researchers' findings (Demian and Morrice, 2012). For example, number of access points to Moodle during the course, first access to standard materials (e.g. lecture slides, notes); timely access to materials (e.g. exercise sheets before relevant lecture); posting to the forum; use of the Storyline revision videos; and interaction with extra-curricular resources above and beyond the course material.

This resulted in 23 potential covariates, with a sample size of 190. Initial scatterplots of final mark against each of the potential covariates revealed very little in terms of a direct an association between any covariate and the outcome. This is to be expected, given the complex relationship between resource usage and achievement, and has been noted elsewhere for other academic disciplines (Demian and Morrice, 2012).

Several automated model building strategies were initially used to reduce the number of covariates, all of which gave nearly identical suggested models. All covariates that were suggested in at least one automated model strategy were used as a base model. This consisted of eight covariates, which was eventually reduced to six covariates via an F-test. The remaining covariates demonstrate that diligence pays dividends in terms of achievement, though intriguingly covariates such as accessing slides and notes in a timely manner are not included.

Some issues were detected with normality of error term, though squaring the response resolved the problem. The square of the overall mark was standardized, yielding estimates for the regression model as given in Table 1. The coefficient of determination for this model was approximately 24%.

Figure 1. Access to Moodle and lecture recordings, from the start of the course until the final exam. Vertical green lines denote the date of the in-class test in November and the final exam in May.
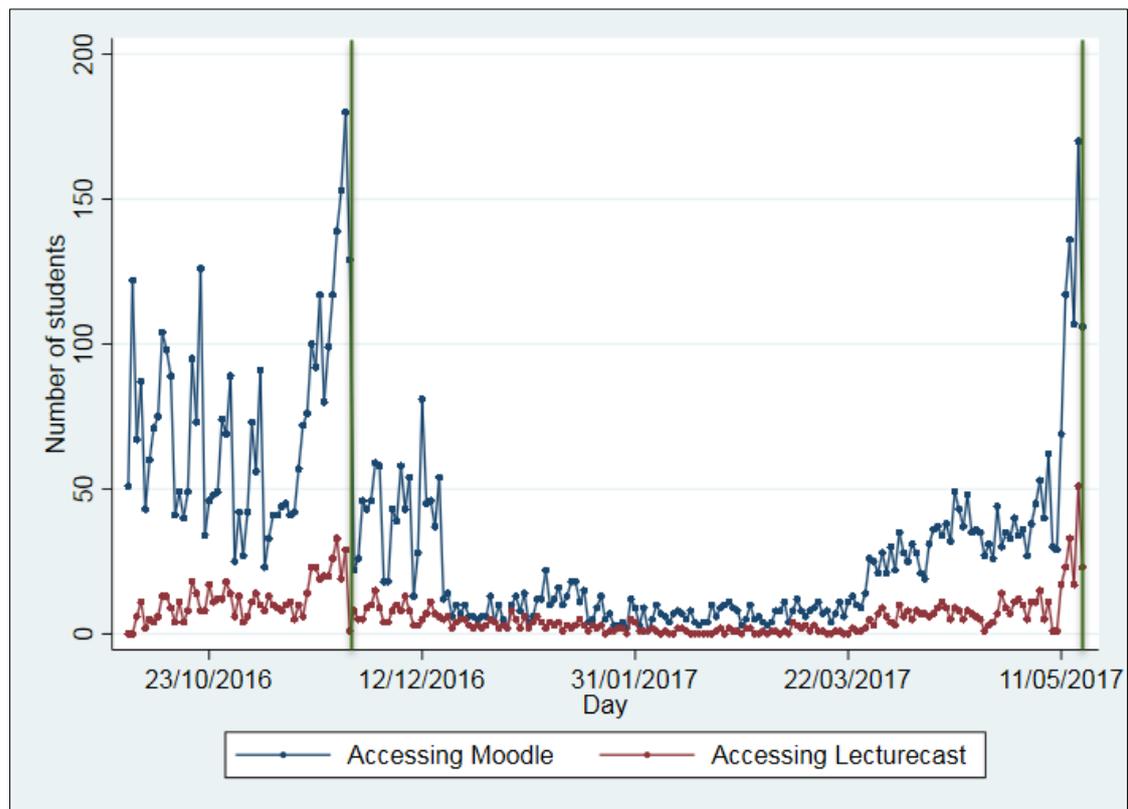
DISCUSSION

Students are overwhelmingly positive about the availability of course resources via a VLE, including access to lecture recordings as it allows complete flexibility in choosing how and when to learn. Though we cannot assess the quality of this learning, or any learning that occurs outside of the VLE, our investigation does allow insight into the quantity of learning activities that students participate in and their general engagement with the course (You, 2016). With growing class sizes, and fewer compulsory requirements, monitoring student participation is a difficult task. Tracking activity on a VLE gives some assurance, in the broadest terms, of engagement with the course.

Despite this flexibility, the results unsurprisingly show that students are driven by assessment rather than steady learning, with usage peaking just before an assessment. No differences were detected in usage of resources available on the VLE, or attainment levels at the end of the course, for students from Statistical Science, Mathematics and other sciences, with the exception of access to multimedia resources: while Mathematics students were more likely to use lecture recordings more heavily, they were significantly less likely to use the revision Storylines. There were no known lecture clashes which might encourage Mathematics students to make more use of the lecture recordings, but it is known that these students have less access to such resources in other modules. Storyline technology was new to the vast majority, if not all students; why Mathematics students were less likely to use this resource is not clear.

Modelling achievement suggested that on average, the more diligent a student in interacting with the resources on a VLE, the better they perform. Though not significant, it is intriguing that those who watched both Storyline revision videos performed slightly better than those who watched neither, but not as well as those who watched only one. This may be a sign of student confidence – that they are already comfortable with the course material. Equally, students

may work together on these interactive videos so that the VLE records are not a true reflection of student engagement.

Table 1. Final regression model for achievement. Note that all covariates are binary, with the exception of the 'average percentage of recorded lecture watched in a single sitting'.

COVARIATE	ESTIMATE (STD.ERROR)	P-VALUE
Intercept	1.046 (0.231)	<0.001
Average percentage of recorded lecture watched in a single sitting	0.005 (0.002)	0.025
Optional exercise sheets: • 3-4 exercise sheets accessed before relevant tutorial (binary variable) • 5-6 exercise sheets accessed before relevant tutorial (binary variable)	0.243 (0.147) 0.617 (0.203)	0.100 0.003
Posted to forum: • During term time (binary variable) • After term time (binary variable)	0.429 (0.202) 0.471 (0.169)	0.035 0.006
Accessed all exam solutions (binary variable)	0.430 (0.137)	0.002
One out of two revision Storylines viewed	0.628 (0.220)	0.005
Two out of two revision Storylines viewed	0.274 (0.222)	0.218

We are limited by the data that a VLE tracks, have no information on individual student attendance at lectures, and have no way of assessing how much independent learning happens outside of the VLE. The latter is likely to have a major impact on achievement, and so a model for attainment based purely on VLE analytics is unlikely to ever provide powerful predictors.

Though it is likely that a change in instructor, in the nature of resources available on the VLE, or the subject matter, would change the way students interact with the material (Demain and Morrice, 2012; Chowdhry, Sieler and Alwis, 2014), these findings mirror those of other research: that interaction with VLE resources is only weakly related to achievement (Demain and Morrice, 2012; Chowdhry, Sieler and Alwis, 2014), and that covariates similar to those chosen for the model here have been found to be related to achievement elsewhere, with similar levels of variability in marks explained by these covariates (Macfadyen and Dawson, 2010).

REFERENCES

Baepler, P., & Murdoch, C. (2010). Academic Analytics and Data Mining in Higher Education. *International Journal for the Scholarship of Teaching and Learning*, *4*(2), 1-9.

Chowdhry, S., Sieler, K., & Alwis, L. (2014) A Study of the Impact of Technology-Enhanced Learning on Student Academic Performance. *Journal of Perspectives in Applied Academic Practice*, *2*(3), 3-15.

Demian, P., & Morrice, J. (2012). The Use of Virtual Learning Environments and their Impact on Academic Performance. *Engineering Education*, *7*(1), 11-19.

Macfadyen, L., & Dawson, S. (2010). Mining LMS Data to Develop an "Early Warning System" for Educators: A Proof of Concept. *Computers and Education*, *54*, 588-599.

You, J. (2016). Identifying Significant Indicators Using LMS Data to Predict Course Achievement in Online Learning. *Internet and Higher Education*, *29*, 23-30.