ASSOCIATION
FOR LEARNING
TECHNOLOGY ALT

ORIGINAL RESEARCH ARTICLE

What can we learn from learning analytics? A case study based on an analysis of student use of video recordings

Moira Sarsfield[a]* and John Conway[b]

[a]*Faculty of Natural Sciences, Imperial College, London, UK;* [b] *Kaleidoscope Health and Care, London, UK*

Over recent years the use of lecture capture technology has become widespread in higher education. However, clear evidence of the learning benefits of this technology is limited, with contradictory findings reported in the literature. The reasons for this lack of consistent evidence may include methodological issues and differences in the context of previous studies. This paper describes a study using server log data to explore student use of video recordings quantitatively in the context of science courses at Imperial College London. The study had two aims: to understand more about the general principles that underpin a learning analytics study and to seek answers to the following specific research questions: (1) How much use is made of video recordings? (2) How does the use of recordings in a module vary over time? (3) Is the use of recordings different for different modules or subjects? (4) Is the use of recordings different for subgroups of students, for example, students with specific learning differences or English as a second language, students attaining different grades? (5) Is the use of recordings different for different types of content? Using learning analytics enabled the discovery of context-specific actionable insights: recommendations for both staff and students and ideas for further research. General conclusions were also drawn on how best to undertake learning analytics studies in order to deliver evidence and insights to improve learning and teaching.

Keywords: learning analytics; video recordings; big data; visualisation; lecture capture

Introduction

Technology to record lectures (lecture capture, LC) and to make other types of video recordings available to students to support learning is now widely used in educational institutions. LC is very popular with students, but some teaching staff have concerns that the technology may have unexpected or even detrimental effects on student learning (O'Callaghan *et al.* 2017; Witthaus and Robinson 2015). Staff are especially interested to learn more about how student use LC recordings. For example, do students 'dip in' to clarify particular points in a lecture – either aspects that were not understood or where concentration lapsed – or do they view entire LC recordings? Further, does provision of LC lead to changes in study patterns, for example lower attendance at lectures, procrastination in writing up notes or placing too much emphasis on the

---

*Corresponding author. Email: m.sarsfield@imperial.ac.uk

1

lecture itself to the detriment of engagement with other learning materials such as textbooks and primary literature? Finally, does the introduction of LC have any impact on lecture attendance and student attainment?

A number of literature reviews on the use of LC were consulted (Deal 2007; Heilesen 2010; Karnad 2013; Witthaus and Robinson 2015). All of the reviews highlight high student satisfaction with LC (which may also be termed *lecture recording*, *podcasts* or *vodcasts*); suggested benefits include increased flexibility for studying and reduced student anxiety. However, the reported proportion of students who use LC when it is available varies considerably; for example Witthaus and Robinson (2015) report values from different studies ranging from around 33% to 96%, with a single study reporting values between 21% and 100% (Turró *et al.* 2014). This wide variation in reported use of LC may arise because, in general, each study reports the value for one specific context, and there are many differences in context and in the research methodologies adopted. Consequently it is difficult to draw conclusions on the reasons for the variation or to compare results between the studies.

The main reasons described for LC use are to clarify points from lectures and for revision; some students also indicate that LC recordings may be used to catch up on missed lectures. Most use is reported to take place shortly after lectures, just before assignments and in the period before examinations. Regarding the impact of LC on attendance and attainment, the studies reported do not provide a clear answer – different studies report positive effects, negative effects or no difference in both attainment and attendance. It is generally held that LC is of benefit to students with specific learning differences or with English as an additional language, based on a small number of studies such as those of Pearce and Scutter (2010) and Leadbeater *et al.* (2013).

The fact there is little consistency between the results observed in different studies is not unexpected because each study is carried out in a particular context, with many different variables at play, such as the characteristics of the students, the subject area of the course, the pattern of teaching and assessment, and the implementation of LC in the institution. Even within a single institution Turró *et al.* (2014) reported wide variation in the use of LC recordings by students in different subjects. The importance of context was also highlighted by Gašević *et al.* (2016) in relation to predictive learning analytics; they highlight the example of the Finnegan, Morris and Lee (2009) study, which described significant differences in the behaviour of successful students on online courses in different academic areas (English, Social Sciences and Science, Technology, Engineering and Mathematics (STEM) subjects). We should thus be wary about generalising specific findings from the literature or suggesting that these may apply in another context.

Karnad (2013) and Witthaus and Robinson (2015) highlight that studies into LC use are often based on self-reporting by students and that responses are often limited to a self-selected sample of students, who may be those who find LC particularly useful. The results of studies that use self-reporting should therefore be treated with caution. Gorissen, Van Bruggen and Jochems (2012a, 2012b) explored the reliability of self-reported data by conducting triangulation of self-reported data on LC use with server log data for the same students. The authors found significant differences and concluded, 'Given the discrepancies between verbal reports and actual usage, research should no longer rely on verbal reports alone'.

Now that LC technology is installed widely and server log data is more readily available, it is increasingly possible to analyse the use of LC using quantitative techniques and to include all students in a study, thus eliminating possible reporting and

sampling bias. Institutions can explore LC use in their own context, exploring differences in use (by subgroups of students, over time, in different subject areas, etc.), and thus build up a more nuanced understanding of how LC recordings are used and draw out context-specific recommendations for best practice.

Aims and objectives

Our aim was to adopt a learning analytics approach to explore how students use video recordings in the context of science courses at Imperial College London, with two objectives: to gain insights into student use of LC technology and also to understand more about the general principles that underpin a learning analytics study. The Higher Education Academy (HEA 2015) defines *learning analytics* as 'the process of measuring and collecting data about learners and learning with the aim of improving teaching and learning practice'. Learning analytics may also be used to make predictions on individual student success or retention, as described by Jisc (Sclater and Mullan 2017):

> Learning analytics systems enable universities to track individual student engagement, attainment and progression in near-real time, flagging any potential issues to tutors or support staff. They can then receive the earliest possible alerts of students at risk of dropping out or under-achieving. (p. 6)

This type of predictive analysis relating to individual students was not the aim of our study. We sought, in line with the HEA definition, to investigate student use of video recordings in order to discover 'actionable insights' (Cooper 2012) on how to improve teaching and learning using this technology and additionally to learn more about the process of using learning analytics techniques, that is, to investigate how best to implement these techniques more widely.

A key consideration was to ensure that the study was conducted with due consideration for ethics and privacy. The guidance on conducting educational research available at the time of the study largely focused on gathering new data to address research questions, whereas we intended to use data that had already been gathered as part of normal business. Emerging best practice guidelines on the use of learning analytics (Jisc 2015; Open University 2015) proved to be very useful in determining the approach we should follow.

The research questions to be investigated were discussed and agreed in advance with senior teaching staff from each department, as follows:

- How much use is made of video recordings (LC and other types)?
- How does the use of recordings in a module vary over time?
- Is the use of recordings different for different modules or subjects?
- Is the use of recordings different for subgroups of students, for example, students with specific learning differences or English as a second language, students attaining different grades?
- Is the use of recordings different for different types of content, for example, recordings of lectures, flipped lectures, post-lecture summaries?

Only the data required to answer these specific research questions were recorded and the data were anonymised in line with the recommendations of the Information

Commissioner's Office (2012). We also followed internal college policy, which is in line with Higher Education Funding Council for England (HEFCE) (2015) practice, that data should be excluded from published reports if the number of students in a particular category is fewer than 10, in order to prevent identification, that is, data were not reported in ways that were attributable to individuals so that anonymity was assured. Guidance from the college's Legal Services Office at the time of the study noted that on entry to the college students agreed to their data being used for 'administrative purposes' and were informed that student data might be used for 'research and statistical analysis'. This indicated that sufficient agreement had been given for our study, and thus all students on each module were included; there was no self-selection of participants, which can be a drawback in studies where specific consent is required (Brooks *et al.* 2014). The new data protection legislation, the *General Data Protection Regulation* (GDPR), which came into effect in EU member states in May 2018, may have implications for the conduct of future similar studies.

The impact of providing LC recordings on student attendance at lectures is of concern to academic staff and is frequently cited as a reason against the use of LC technology (O'Callaghan *et al.* 2017; Witthaus and Robinson 2015). However, it was decided not to investigate attendance in this study, for a number of reasons. Firstly, lecture attendance was not mandatory on any of the modules studied, and so a register was not customarily taken. Gathering data on lecture attendance for the purposes of the study would have required consent from students, which would almost certainly have reduced the number of participants and introduced sampling error. Secondly, taking a register in lectures could potentially have caused anxiety for students and changes in their behaviour. Finally, no baseline data were available for comparison.

Methodology

Study design and technical details

In designing the study, we drew upon the principles outlined by Miller and Mork (2013), who discuss the 'value chain' of data within an organisation and highlight how data from different sources can be brought together to provide insight and information to inform decision-making. This involves 'data discovery', including consideration of ownership and access; 'data integration', where data are brought into a common format, to allow comparisons to be made; and 'data exploitation', which includes analysis, visualisation and examination to determine actionable insights. A similar process is suggested by Gorissen *et al.* (2012b), who also emphasise the importance of data cleaning, and by Jagadish *et al.* (2014), who discuss the importance of feedback and validation at each stage to ensure that the data are correct and consistent and can safely be used in comparisons. We therefore built data cleaning and validation into the study design, as discussed in more detail later.

Technically, the study used Microsoft Excel for data gathering and initial processing, because it is a familiar product for academic and administrative staff and the files are easy to share, save and distribute. Specialist data analysis tools – R software (R Core Team 2017) and R Studio (RStudio Team 2015) – were used for the detailed analysis, data visualisation and reporting, enabling automated scripting of these processes. The overall workflow adopted for the study is summarised diagrammatically in Figure 1.
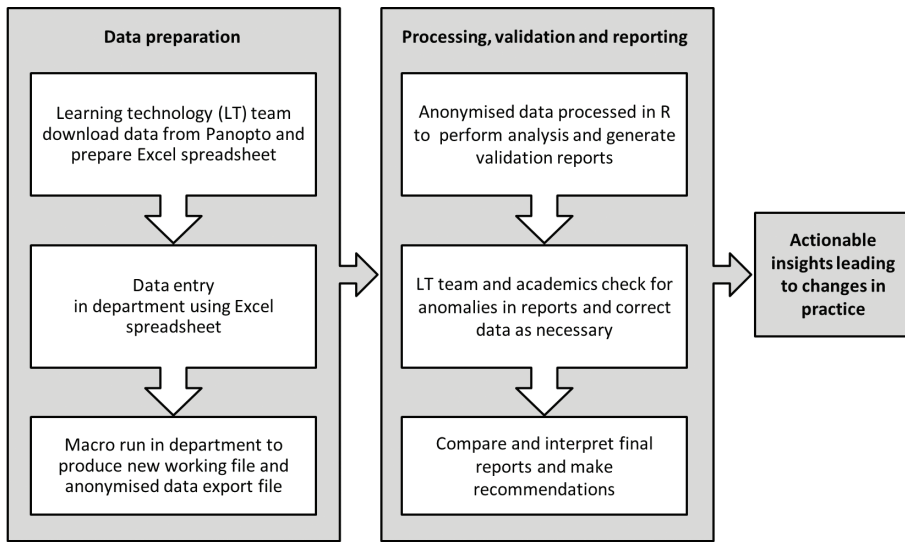
Figure 1. Outline of the workflow of the study.

Data preparation

The study investigated the use of video recordings, made using the Panopto recording system, on 17 undergraduate modules from years 1 and 2 of degree programmes in Biochemistry, Biology, Chemistry, Mathematics and Physics for the academic year 2014–2015. A module was defined as a single block of teaching ending in an examination, and all were taught face-to-face on campus. The modules to be investigated were selected by the departments as being representative of their degree programme; large classes were selected so that generally sufficient numbers of students were included in each subgroup to avoid privacy concerns.

The data collected and used in the study are described in Table 1, and the types of recording investigated are listed in Table 2. In some departments LC recordings were automatically scheduled and thus were listed by Panopto with a standard duration of 50 min. When a scheduled lecture ended earlier than the listed end time, department staff amended the duration in the data to the actual value. The study only included students who took the module and examination for the first time in 2014–2015, that is, repeating students and students who did not take the final examination at the end of the module were excluded from the study, because their use of the recordings was likely to be atypical.

In the departments no single member of staff had access to all the necessary data, and so the workflow allowed data to be added incrementally. The information on specific learning difference status (which indicated students with dyslexia, dyspraxia, etc.) could be considered sensitive under the terms of the *Data Protection Act* and so, to maintain confidentiality, this was added last. The data were then processed using an Excel macro that produced two time-stamped output files: an updated spreadsheet including all data, which was retained securely in the department and could be checked, revised and reprocessed, as required; and a data export file in comma-separated values (CSV) format, in which identifying student data were anonymised by replacing student usernames with a hashed value. The latter file was passed to the learning technology team for further processing.

Table 1. Data used in the study.

Record	Field	Source
Recording	Name, start date and time, end date and time, module	Panopto
	Duration	Calculated/Department
	Type	Department
User access	Username, recording name, access date and time, minutes viewed	Panopto
Student	Username, specific learning difference status, fee status, grade attained	Department
Module	Department, year, core or optional, date of examination	Department

Table 2. The types of recording investigated in the study.

Type	Description
Lecture	A traditional lecture delivered live to students.
Lecture-Faulty	A faulty recording of a lecture, not of use to students, usually because of problems with audio capture.
Lecture-NonExamined	A lecture consisting entirely of material that will not be examined.
PreSession	Recordings that are designed to be viewed by students before a teaching session (lecture, practical, tutorial), for example, when flipping the classroom or pre-sessional materials for team-based learning.
OfficeHours	A session where problems are worked through with individual students or a small group.
Other	Anything other than a Lecture or Office Hours that is delivered directly to students, for example, a problem class, tutorial or practical session.
PostSession	Recordings made after a lecture, practical or tutorial, providing further information for the students.

Data processing, validation and reporting

Use of recordings can be investigated in two ways: by number of accesses or by minutes viewed. The latter measure provides more detailed information, but additional processing is needed. Users can access a recording either by viewing it in the Panopto viewer, in which case minutes viewed is logged in the user access data, or by downloading the recording to a local device, which results in an access of zero minutes' duration being logged. An 'adjusted minutes viewed' value was calculated, as described in Appendix 1, to allow for viewing of downloaded recordings. To summarise the use of recordings over time, accesses were allocated into two time periods, designated the 'learning period' and the 'revision period'. The definitions of these terms are detailed in Appendix 1.

The use of LC recordings was examined in subgroups of students categorised by attainment on the module, specific learning difference status and fee status. Fee status was used as a proxy for students with English as an additional language, although there is not a direct correspondence and so this is only an approximation. Other factors analysed were use of non-lecture recordings and timing of use in relation to the

timetable for the module for all students and for subgroups of students categorised by attainment on the module.

Validation reports were prepared for each of the modules and returned to the departments for scrutiny. These reports included tables of data and visualisations designed to highlight particular aspects of the data for checking purposes and thus enabled problems in the data to be picked up and corrected. For example, it was possible to identify recordings that were faulty or miscategorised or to detect instances where the length of a recording was incorrectly reported, as illustrated in Figure 2. In this case the final four lectures (L21-L24) show considerably lower use than the other lectures in the module. On accessing these lecture recordings it was found that the audio track was missing because of a technical fault. These four lectures were therefore redesignated as 'Lecture-Faulty', and thus the data relating to the recordings were not included in the LC analysis. This image also highlights the importance of including the correct duration for each recording, which can vary considerably. In this case the correct information has already been recorded, as shown by the variation in duration represented by the large grey dots. Omitting this step would result in incorrect values for percentage viewed values. Correction of these issues ensured that the data were consistently reported, and so it was safe to make comparisons between modules. Any changes to the data at this stage were made in the department and the Excel macro was run again, producing new files for further processing.

Standard final reports were prepared for all the modules studied, each including the same analyses to allow comparisons to be made between modules and subjects. The final reports were reviewed by the module convenors and senior teaching staff from each department in order to draw out insights from the analyses.

The methodology adopted for this study is suitable for occasional use, but not for regular, real-time reporting. This is because the analyses are run as a batch process, data gathering is manual and cumbersome, and many of the analyses rely on having access to the full dataset, including examination results and timing of use across
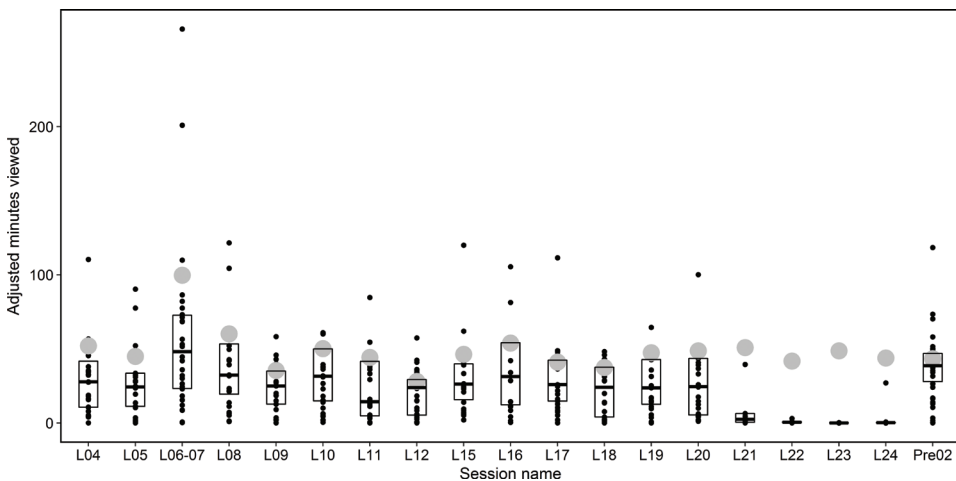

Figure 2. Number of minutes viewed for each recording on the Life Sciences module Behavioural Ecology. Each black dot represents the adjusted minutes viewed for a single student, and the large grey dot represents the duration of the recording. The median and interquartile range is also included.

the whole study period. Following discussions with teaching staff, training materials were developed to show how built-in functionality in the Panopto system can provide them with real-time feedback on which parts of recordings are viewed most frequently (these may be specific topics that students find particularly difficult to understand) and also on overall volume and timing of viewing for individual recordings and for all recordings in a module. This functionality provides an interim way for staff to monitor use of recordings, until a production-scale learning analytics system is introduced.

Results and discussion

Overall use of LC recordings

LC recordings were widely used across all modules studied, as summarised in Table 3. However, use varied considerably between modules, with the percentage of students viewing at least one LC recording on a particular module ranging from 26% to 98%. This is similar to the range of values reported in the literature; for example Witthaus and Robinson (2015) reported on a number of studies, with values ranging from 33% to 96%. The average percentage of all recorded minutes viewed ranged from 3% to 36% on different modules. This may seem to be in line with students dipping in to view small segments of recordings, but in fact there was considerable variation in use, with

Table 3. Overall use of lecture capture recordings across all modules.

Subject and module	Year and module type	Percentage of students who viewed LCs	Average (maximum) percentage of all recorded minutes viewed	Percentage of initial accesses in the learning (revision) period
Biochemistry				
Genes and genomics	Y2, core	83	24 (171)	36 (64)
Macromolecular Structure and Function	Y2, core	93	34 (143)	48 (52)
Biology				
Applied Molecular Biology	Y2, core	88	29 (168)	56 (44)
Behavioural Ecology	Y2, optional	78	30 (149)	35 (65)
Virology	Y2, optional	96	36 (102)	36 (64)
Chemistry				
Heteroaromatic Chemistry	Y2, core	67	19 (114)	76 (24)
Organic Synthesis 1	Y2, core	57	16 (107)	71 (29)
Organic Synthesis 2	Y2, core	70	18 (124)	74 (26)
Quantum Chemistry	Y2, core	67	15 (103)	70 (30)
Mathematics				
Analysis	Y1, core	98	12 (56)	82 (18)
Mathematical Methods	Y1, core	96	12 (89)	90 (10)
Statistics	Y1, core	84	8 (62)	94 (6)
Differential Equations	Y2, core	78	10 (74)	84 (16)
Physics				
Mechanics	Y1, core	26	10 (84)	62 (38)
Vibrations and Waves	Y1, core	36	18 (117)	53 (47)
Statistics	Y2, core	54	11 (89)	49 (51)
Thermodynamics	Y2, core	72	3 (71)	57 (43)

LC, lecture capture.

some students viewing no LCs, while others viewed over 100% of the available LC minutes (because some material was viewed more than once). This demonstrates the value of recording data for each individual student rather than drawing conclusions from an average value.

In general, similar results were observed for modules within a subject, but clear differences were seen between subjects. For example, the average percentage minutes viewed was consistently higher for modules in Life Sciences (Biochemistry and Biology) compared with modules in Physical Sciences (Chemistry, Mathematics and Physics). This reflects the fact that many views in Life Sciences were of whole recordings, whereas in Physical Sciences a much larger proportion of views could be categorised as dipping in, that is, viewing a short, specific section (Figure 3). This is particularly the case for Mathematics modules, which show low values of average percentage of recorded minutes viewed in spite of the high percentage of students who viewed recordings. This may align with the findings of Cortinhas (2017), who studied students taking economics modules at the University of Exeter and reported that the volume of use of LC recordings was significantly lower on 'quantitative modules' (with a mean value of 1.709 h per term vs. 3.206 h per term for 'non-quantitative modules'). However, it is not clear how similar quantitative modules, as defined by Cortinhas, are to the Mathematics modules at Imperial College.

On Physical Sciences modules most LC use occurred early in the module, just after the lecture was delivered, with only a small volume of additional use in the period before the examination. The pattern of access for Life Sciences LC recordings was very different, with a much higher proportion of initial accesses (up to 65%) occurring in the revision period rather than in the learning period. The prevalent pattern of use of LCs by Mathematics students is therefore to dip into the recordings shortly after the lecture is delivered and then not to return to the recording again. This pattern may result from the design of Mathematics modules, in which knowledge and techniques are introduced sequentially, with regular testing of understanding. Once a topic is
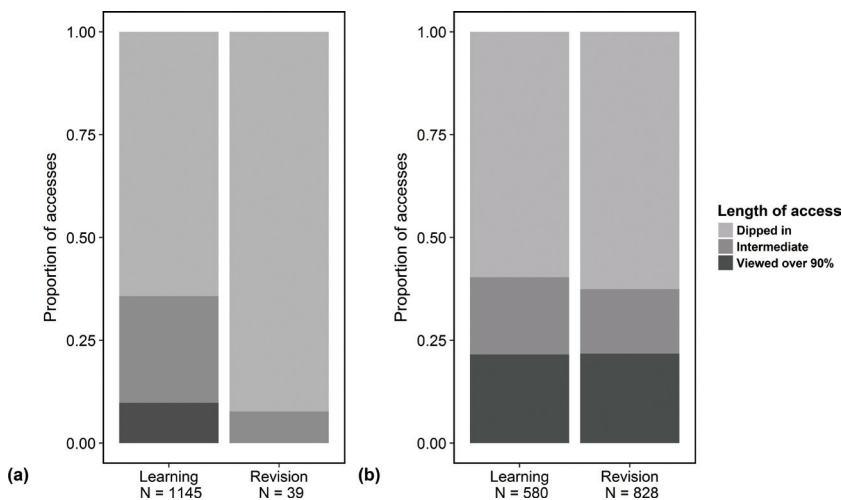

Figure 3. For lecture capture recordings on (a) the Statistics module and (b) the Applied Molecular Biology module, the proportion of accesses of different length in the learning period and revision period (raw data only). The categories represent the duration of each access relative to the duration of the entire recording, as follows: 'dipped in', <30%; 'intermediate', 30%–90%; 'viewed over 90%', >90%.
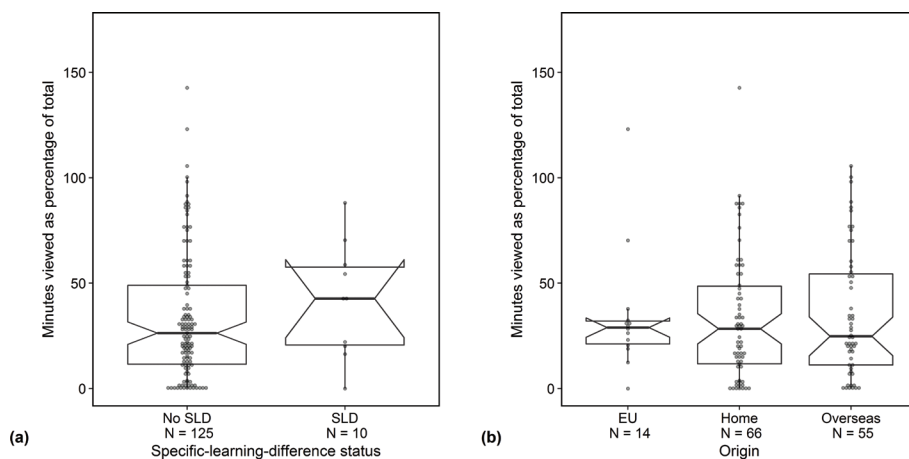
Figure 4. Percentage of all minutes of lecture capture recordings viewed by students categorised by (a) specific learning difference status and (b) fee status on the module Macromolecular Structure and Function. The boxplots show the median and interquartile range, with notches indicating 95% confidence intervals. Individual data points are superimposed.

mastered, there is usually no need to view the lecture content again. In Life Sciences, topics are more wide-ranging, and more synthesis is required, which may explain the continuing high use of recordings throughout the module.

Use of LC recordings by subgroups of students

For subgroups of students categorised by specific learning difference status and by fee status, no clear evidence of difference in the use of LC recordings was observed between subgroups on any of the modules studied, as illustrated in Figure 4, which shows the results of this analysis for one module. Therefore, although other studies in other contexts have reported increased use by these subgroups (e.g. Cortinhas 2017; Leadbeater *et al*. 2013; Pearce and Scutter 2010), within the context of science courses at Imperial, there was no evidence that LC recordings were used more by students with specific learning differences or by those originating from outside the UK.

Examining the use of LC recordings by subgroups of students categorised by grade attained also showed no clear difference between the subgroups; that is, no general correlation was observed between the use of LC recordings and attainment on any of the modules studied. An example of the output is shown in Figure 5.

However, some differences were observed in the pattern of use between subgroups of students categorised by grade. Figure 6 shows a lecture-by-lecture view of the average percentage of each LC recording viewed by students achieving different grades on a particular module. Lecture 15 (L15) shows notably high use by students who went on to achieve a first class grade. The lecturer observed that this lecture includes particularly difficult content. Use of recordings by students who went on to fail the module drops noticeably after this lecture, while better-performing students continue to use the LC recordings for the remainder of the module. This finding, and similar observations from other modules, resulted in explicit advice for students, as discussed in the section 'Actionable insights discovered'.
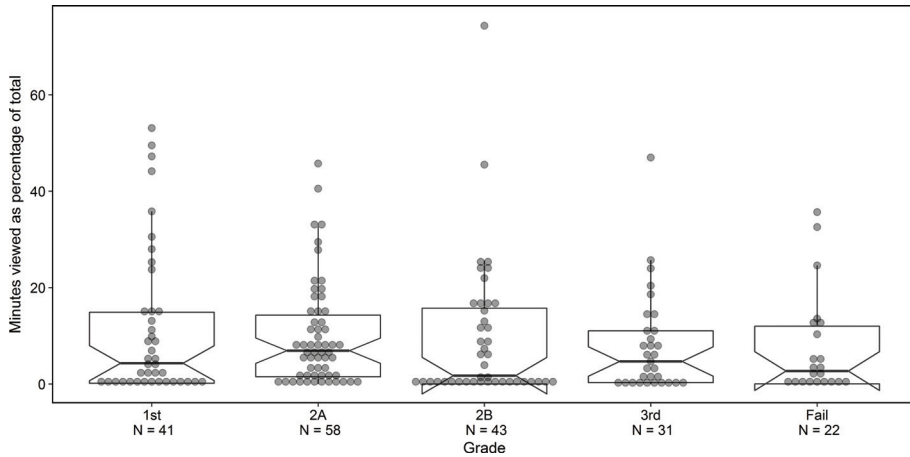
Figure 5. Percentage of all minutes of lecture capture recordings viewed by students categorised by grade attained on the Differential Equations module.
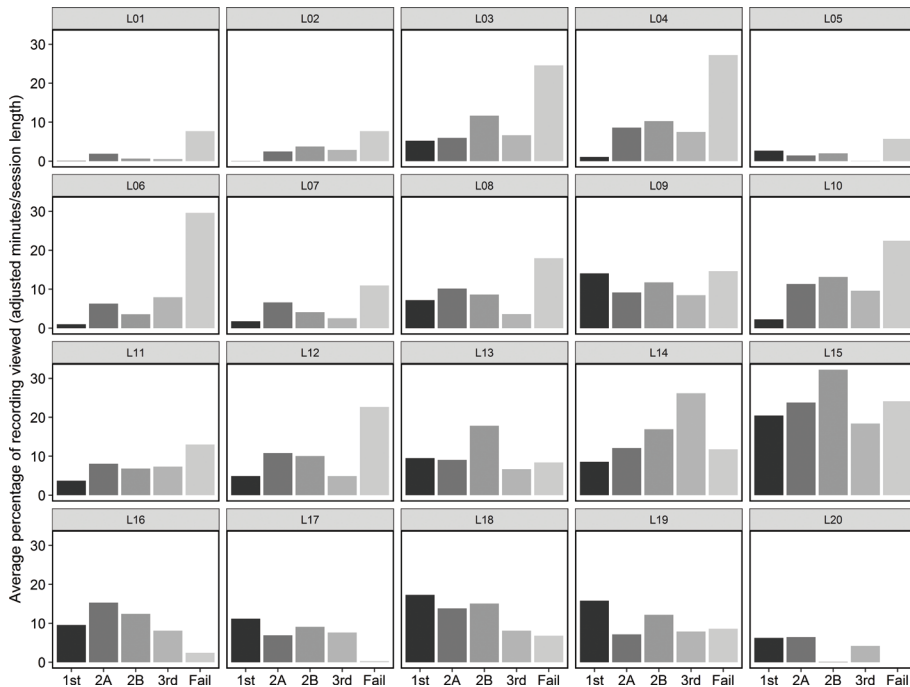

Figure 6. Percentage of each lecture capture recording on the Statistics module viewed on average by students categorised by grade. Number of students in each grade category: 1st, 75; 2A, 56; 2B, 45; 3rd, 25; Fail, 13.

A correlation was also observed between the final grade attained and the timing of initial use of LC recordings, as shown in Figure 7. Students who attained higher grades accessed the recordings more during the learning period, whereas students who attained lower grades tended to access the recordings later, during the revision period. This accords with the findings of Brooks *et al.* (2014), who analysed use of LC recordings
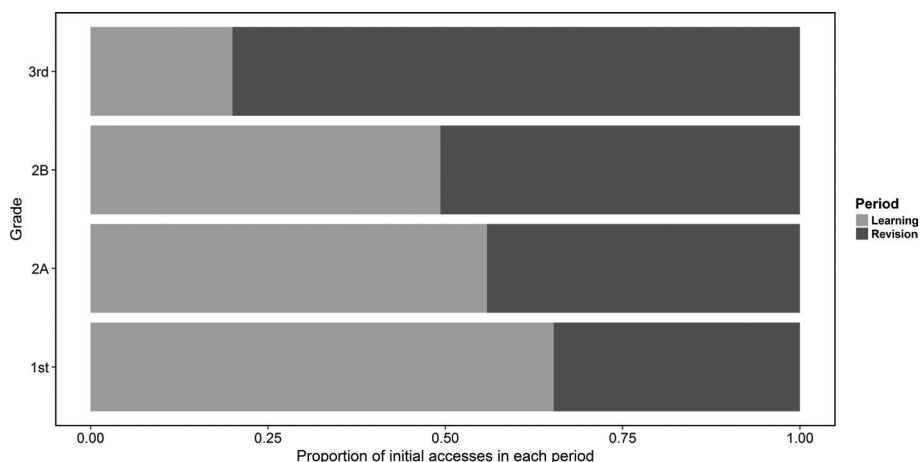
Figure 7. Proportion of initial accesses to lecture capture recordings on the Applied Molecular Biology module that fell within the learning period and revision period for students categorised by grade. Number of students in each grade category: 1st, 28; 2A, 71; 2B, 21; 3rd, 3.

using *k*-means clustering. They identified five patterns of activity among users of LC recordings; students adopting the pattern they labelled 'high activity' (regular use of the recordings throughout) showed better performance in assessment, with average marks ranging from 9.18% to 16.45% higher than the marks attained by students adopting other patterns of use, such as studying only in the period before examinations. Chai (2014) also studied the timing of viewing of LC recordings and reported that 'online lecture recordings are only positively correlated with academic achievement if used during the non-binge study period' (i.e. during the learning period rather than the revision period). Again conclusions can only reliably be related to the specific context of each study; different successful patterns of LC use are likely to be observed for different patterns of assessment and for students studying different academic subjects.

Use of other types of recordings

Other types of recordings were generally used considerably less than LC recordings. For example, on the Applied Molecular Biology module, the average percentages of the cohort viewing recordings designated as 'Lecture-NonExamined' and 'Other' were 5.3% and 13.2%, respectively. However, where a recording was 'required viewing', higher values were observed, with 77.6% of the cohort viewing 'PreSession' recordings that were required viewing prior to a flipped classroom session. For these 'required viewing' recordings, a particular pattern of use was observed (as illustrated in Figure 8), in which the average percentage of the recording viewed was correlated with the grade attained, with students who attained higher grades viewing more of the recording on average than students who attained lower grades. In this case none of the students who went on to fail the module viewed the recording at all. No correlation with grade was observed for any other recording type.

Additional insights from the data

In addition to revealing information relating to the original research questions posed, the data revealed a number of unexpected findings. For example, clear evidence was
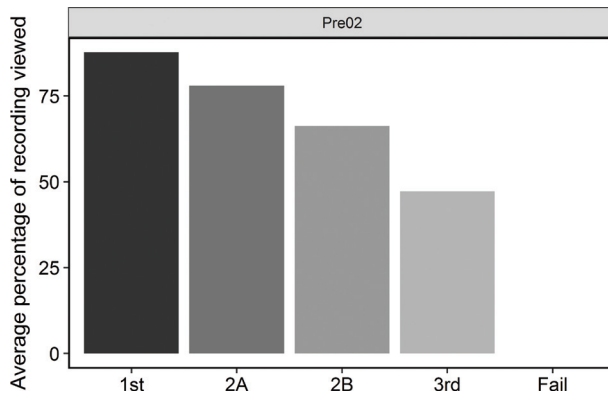
Figure 8. For a 'required viewing' recording on the Behavioural Ecology module, the percentage of the recording viewed on average by students categorised by grade. Number of students in each grade category: 1st, 7; 2A, 33; 2B, 16; 3rd, 2; Fail, 2.
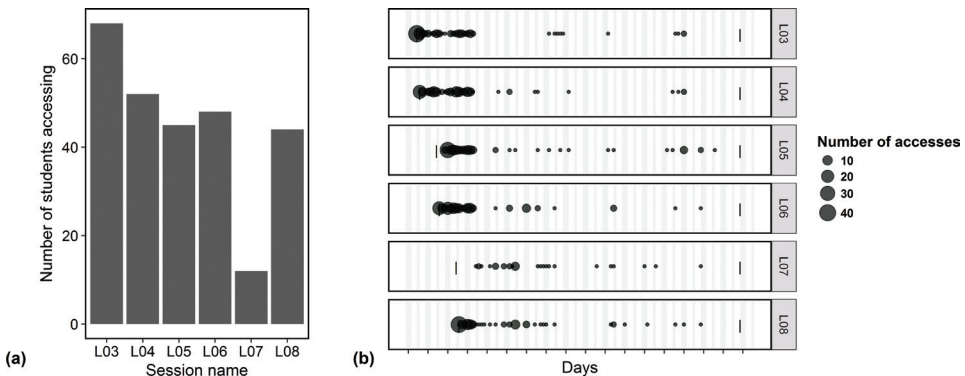

Figure 9. The impact of late release of a lecture capture recording on overall use. Lecture 7 (L07) on the Differential Equations module took place on 04 February 2015, but was not released in Panopto until 09 February 2015. The overall use is considerably less than other lectures on the module. (a) The number of accesses for each recording. (b) Distribution of individual accesses over time. The date of each lecture is shown with a vertical bar on the left, and the examination date is shown with a vertical bar on the right. Weekends and Wednesdays are shaded.

seen that when recordings were released late (as a result of technical or administrative issues), they were accessed less in total than those recordings that were released immediately after the lecture. An example of this is shown in Figure 9, and a similar pattern was observed on several other modules where a LC recording was released late.

For some modules, examination of use on a lecture-by-lecture basis, as in Figure 9a, showed a particular pattern of use that related to timetabling. For example on one Chemistry module higher use was consistently seen for recordings of lectures timetabled for 09:00 on a Thursday morning. The lecturer reported that attendance was usually poor in this time slot and the log data suggest that the recordings were frequently used by students to catch up on these missed lectures prior to the next timetabled lecture in the module.

Another unexpected insight concerned the timing of use of LC recording when the lectures followed different timetabling patterns – either spaced or blocked. Many modules on the Mathematics degree programmes follow a pattern of two lectures per week, with time between these for consolidation of understanding. The use of the recordings of these spaced lectures (see Figure 10a) showed that students generally viewed the recordings promptly after the lecture and then did not return to the recording subsequently. In contrast many modules in Life Sciences include blocks of lectures, sometimes with two or three lectures timetabled in 1 day. Covering the theoretical content in a short period at the beginning of the module allows time for extended practical classes to run in subsequent weeks. A typical pattern of use in Life Sciences (see Figure 10b) shows significant access to recordings over many days. This may be because the blocked pattern of teaching does not allow students to 'catch up' with content from one lecture before the beginning of the next. Thus students may have difficulty following subsequent lectures if they have not yet fully understood particular concepts from an earlier lecture. This may account, at least in part, for the relatively greater use of LC recordings observed on Life Sciences modules.

Actionable insights discovered

The use of learning analytics to study student use of video recordings uncovered a number of 'actionable insights' – ways for both students and academic staff to change or improve existing practice. Actionable insights for lecturers and module or degree organisers were derived from the study and subsequent discussions, as follows. This advice has been disseminated to teaching staff and incorporated into staff training.
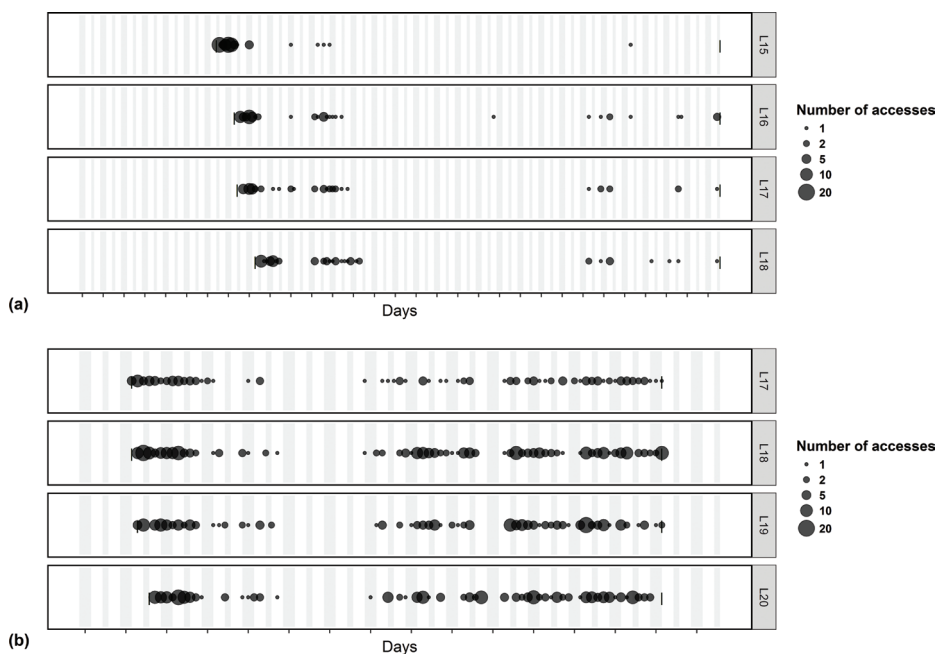

Figure 10. The timing of accesses to lecture capture recordings on (a) a Mathematics module, Statistics; and (b) a Life Sciences module, Macromolecular Structure and Function.

- Do not delay the release of recordings; delayed release results in lower usage (as shown in Figure 9).
- Consider how lectures are timetabled. Students may require time between lectures to assimilate complex content, as illustrated by the different patterns of viewing in Figure 3, Figure 10 and Table 3 for modules on the Mathematics and Life Sciences degrees.
- Use the inbuilt functionality of the Panopto system to view overall volume and timing of viewing for individual recordings and for all recordings in a module and to check which parts of recordings are viewed most frequently; these may be specific topics that students found particularly difficult to understand.
- If the pattern of use of recordings is not as expected, investigate why this is so and make changes as necessary, for example, looking at timing of lectures, volume of material covered, pattern of assessment and so on.
- Think about how recordings are presented. Should they be given more, less or equal prominence compared with other learning materials?
- Give advice to students on the way you expect them to use LC recordings.

The two final points arose from concerns that the high use of LCs by a small number of students may be a less effective use of their study time, and therefore appropriate guidance on LC use is important.

Advice for students was derived from examining the ways that high-performing students tended to use recordings. Across all the modules studied, high-performing students consistently adopted the following patterns of behaviour:

- They viewed recordings early, that is, immediately after the lecture rather than in the revision period as shown in Figure 7.
- They maintained their level of application throughout the module as shown in Figure 6.
- They viewed recordings when the lecturer said it was required (e.g. a flipped lecture) as shown in Figure 8.

However, high performing students did not use LC recordings more or less than poorer performing students; success was not directly correlated with LC recording viewing (Figure 5). The advice now given in study skills lectures at Imperial College is therefore that use of LC recordings may help with learning and that students should decide whether or not it is useful for them. However, if they do use LC recordings in their studies, they should do so promptly, around the time of the lecture, and this pattern of prompt study should be used for all lectures. Finally, they should view recordings that are highlighted by the lecturer as 'required viewing'.

Further research

As a result of the study, two interesting areas were identified for further investigation. Firstly, it would be useful to explore in more detail the ways that high-performing students use LC recordings and to find out whether this differs from the use made by other students. This may uncover further advice for students on optimal ways to use recordings in their studies. Secondly, there are a number of possible reasons why students in Life Sciences use LC recordings differently to students in Physical Science subjects, for example, because of differences in the subject matter itself; in

timetabling, as mentioned previously; in assessment practices; or other factors not yet identified. An exploration of these factors may uncover actionable insights for staff and/or students.

Qualitative research methods, such as interviews or questionnaires, will be needed to explore these questions. This illustrates the general point that quantitative studies can often provide answers to research questions that begin with 'what', 'when' or 'how much', but they can't explain 'why'. A benefit of undertaking a quantitative study first is that it may highlight specific questions to be investigated further using a qualitative approach. Also, with consent, follow-up qualitative studies can include triangulation against earlier quantitative studies, thus avoiding some of the problems associated with self-reporting.

Conclusions

What can we learn from learning analytics?

The case study reported here illustrates that using learning analytics is a successful technique to uncover 'actionable insights' for staff and students relating to use of video recordings. Using the same process to analyse 'click data' from other online learning systems is likely to result in further useful insights. Careful decision-making is needed on the methodology to be adopted for an analytics study. For example, should the study include all students and use data that has already been collected for normal business purposes, or should new data be gathered relating to specific research questions? If the latter approach is adopted, how can recruitment to the study be maximised and selection bias avoided? It is also important to consider carefully what factors and subgroups should be included in the investigation.

This study demonstrated that even within a single faculty in one institution considerable differences were observed in the way that recordings were used by students, especially between subjects. This aligns with the findings of Turró *et al.* (2014) and Finnegan *et al.* (2009), who observed significant differences in the use of technologies by students studying different subjects. As a result it is recommended that studies should be conducted at a module or subject level. Our findings also corroborate the importance of the learning context, as highlighted by Gašević *et al.* (2016), who observed in relation to differences seen in the results of predictive learning analytics studies, 'The under-explored role of contextual variables may help explain the mixed findings in the field … and plausibly these are located in the distinctive elements of the courses that comprised the studies'. Thus we should be careful about drawing wider conclusions from the specific findings of an individual analytics study, because each study does not necessarily provide insight beyond its own particular context. However, reported studies can be very useful in suggesting factors that merit investigation in other contexts. For example, a recent learning analytics study concerning the impact of attendance and use of LC recordings on attainment (Nordmann *et al.* 2018) highlights that prior attainment and year of study are also important factors to consider.

Careful use of learning analytics techniques is likely to result in continued improvements in understanding of student learning within specific learning contexts.

Lessons for future deployment of learning analytics projects

This study highlights a number of points to note for similar learning analytics projects, both at small scale and for larger, production-scale systems. Firstly, results should be

analysed at a fine-grained level, so that key differences in use can be detected. Analysing use by anonymised individual students, rather than working with summary statistics, allows the rich detail of individual actions to be studied. Timing of use is also an important factor to study, especially when this is linked to the timing of other related activities such as assessments and examinations.

Ethical and privacy issues must be considered for all studies that involve student data. In small-scale studies, where data is gathered specifically to address the study's research questions, it may be relatively straightforward to put appropriate processes and data security measures in place, because the data and the users are clearly defined. However, the situation can be more complex and there may be more possibility of a data breach in a larger, production-scale system, which may use data from a central data store, not specifically designed to capture data for the particular study. Risks to data privacy can be identified using data protection impact assessment (DPIA), which is a systematic process introduced by the GDPR that is designed to identify such risks and to minimise these by the use of appropriate processes and mitigations. These could include, for example, data anonymisation, access permissions and rules for reporting (e.g. aggregating or omitting results for any subgroup for which the number of students is below an agreed threshold). It may be possible to automate some or all of the mitigations in a production system. DPIA should be used when a new system or study is designed and revisited when changes are made.

Raw data should be verified by appropriate staff (generally, academic staff who taught on the module) and corrected or excluded as necessary to ensure that the data provide an accurate record of what happened in reality. The automated production of data tables and visualisations can help in this data verification step. If this stage is omitted it is unsafe to make comparisons or draw conclusions from the raw data. Production-scale learning analytics systems should therefore enable validation and correction of data as a standard feature, with details of all changes being logged and auditable.

The system must enable standardised reports to be run on different datasets, allowing comparisons to be made between modules, subjects and so on. If possible, flexibility in reporting should be included, for example, enabling comparisons between additional subgroups or using assessment measures other than the final results of the module. Ideally in production systems, data collection, validation and reporting can occur while the module is running, providing immediate feedback for teaching staff on student use of resources.

Finally, it is not sufficient just to record and report; the results of the analyses must be interpreted and translated into recommendations and actions that will improve student learning, which is the ultimate aim of the learning analytics process.

Acknowledgements

References

Brooks, C., *et al.*, (2014) 'Modelling and quantifying the behaviours of students in lecture capture environments', *Computers & Education*, vol. 75, pp. 282–292. https://doi.org/10.1016/j.compedu.2014.03.002

Chai, A. (2014) 'Web-enhanced procrastination? How online lecture recordings affect binge study and academic achievement', Discussion Papers in Economics economics: 201404, Griffith University, Department of Accounting, Finance and Economics. Brisbane, pp. 26, [online] Available at: https://www120.secure.griffith.edu.au/research/file/efaabb10-acfc-4658-a291-9789b2111321/1/2014-04-web-enhanced-procrastination-how-online-lecture-recordings-affect-binge-study-and-academic-achievement.pdf

Cooper, A. (2012) 'What is analytics? Definition and essential characteristics', *CETIS Analytics Series*, vol. 1, no. 5, pp. 1–10. [online] Available at: https://pdfs.semanticscholar.org/98ab/3fbde3c583d30adf8e660a30e840ebaf2bf0.pdf

Cortinhas, C. (2017) 'Is lecture capture benefiting (all) HE students? An empirical investigation', (No. 1706), Exeter University, Department of Economics. [online] Available at: http://people.exeter.ac.uk/cc371/RePEc/dpapers/DP1706.pdf

Deal, A. (2007) *Carnegie Mellon Teaching with Technology White Paper: Lecture Webcasting*, [online] Available at: https://www.cmu.edu/teaching/technology/whitepapers/LectureWebcasting_Jan07.pdf

Finnegan, C., Morris, L. V. & Lee, K. (2009) 'Differences by course discipline on student behavior, persistence, and achievement in online courses of undergraduate general education', *Journal of College Student Retention: Research, Theory and Practice*, vol. 10, no. 1, pp. 39–54. https://doi.org/10.2190/cs.10.1.d

Gašević, D., *et al.*, (2016) 'Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success', *The Internet and Higher Education*, vol. 28, pp. 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002

Gorissen, P., Van Bruggen, J. & Jochems, W. (2012a) 'Students and recorded lectures: Survey on current use and demands for higher education', *Research in Learning Technology*, vol. 20, no. 3, pp. 297–311. https://doi.org/10.3402/rlt.v20i0.17299

Gorissen, P., Van Bruggen, J. & Jochems, W. (2012b) 'Usage reporting on recorded lectures using educational data mining', *International Journal of Learning Technology*, vol. 7, no. 1, pp. 23–40. https://doi.org/10.1504/ijlt.2012.046864

HEFCE (2015) *2015 National Student Survey: Publication of Data*, Circular letter 12/2015, [online] Available at: http://www.hefce.ac.uk/pubs/Year/2015/CL,122015/Title,103898,en.html

Heilesen, S. B. (2010) 'What is the academic efficacy of podcasting?', *Computers & Education*, vol. 55, no. 3, pp. 1063–1068. https://doi.org/10.1016/j.compedu.2010.05.002

Higher Education Academy (HEA) (2015) *Learning Analytics*, [online] Available at: https://www.heacademy.ac.uk/knowledge-hub/learning-analytics-0

Information Commissioner's Office (2012) *Anonymisation: Managing Data Protection Risk, Code of Practice*, [online] Available at: http://ico.org.uk/for_organisations/data_protection/topic_guides/~/media/documents/library/Data_Protection/Practical_application/anonymisation-codev2.pdf

Jagadish, H. V., *et al.*, (2014) 'Big data and its technical challenges', *Communications of the ACM*, vol. 57, no. 7, pp. 86–94. https://doi.org/10.1145/2611567

Jisc (2015) *Code of Practice for Learning Analytics*, [online] Available at: https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics

Karnad, A. (2013) *Student Use of Recorded Lectures: A Report Reviewing Recent Research into the Use of Lecture Capture Technology in Higher Education, and Its Impact On Teaching Methods and Attendance*, LSE, London. Available at: http://eprints.lse.ac.uk/50929/1/Karnad_Student_use_recorded_2013_author.pdf

Leadbeater, W., *et al*., (2013) 'Evaluating the use and impact of lecture recording in undergraduates: Evidence for distinct approaches by different groups of students', *Computers & Education*, vol. 61, pp. 185–192. https://doi.org/10.1016/j.compedu.2012.09.011

Miller, H. G. & Mork, P. (2013) 'From data to decisions: A value chain for big data', *IT Professional*, vol. 15, no. 1, pp. 57–59. https://doi.org/10.1109/mitp.2013.11

Nordmann, E., *et al*., (2018) 'Turn up, tune in, don't drop out: The relationship between lecture attendance, use of lecture recordings, and achievement at different levels of study'. *Higher Education*. https://doi.org/10.1007/s10734-018-0320-8.

O'Callaghan, F. V., *et al*., (2017) 'The use of lecture recordings in higher education: A review of institutional, student, and lecturer issues', *Education and Information Technologies*, vol. 22, no. 1, pp. 399–415. https://doi.org/10.1007/s10639-015-9451-z

Open University (2015) *Policy on Ethical use of Student Data for Learning Analytics*, [online] Available at: http://www.open.ac.uk/students/charter/sites/www.open.ac.uk.students.charter/files/files/ethical-use-of-student-data-policy.pdf

Pearce, K. & Scutter, S. (2010) 'Podcasting of health sciences lectures: Benefits for students from a non-English speaking background', *Australasian Journal of Educational Technology*, vol. 26, pp. 1028–1041. https://doi.org/10.14742/ajet.1032

R Core Team (2017) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/

RStudio Team (2015) *RStudio: Integrated Development for R*, RStudio, Inc., Boston, MA. Available at: http://www.rstudio.com/

Sclater, N. & Mullan, J. (2017) *Learning Analytics and Student Success: Assessing the Evidence*, Jisc, Bristol, [online] Available at: http://repository.jisc.ac.uk/6560/1/learning-analytics_and_student_success.pdf

Turró, C., *et al*., (2014) 'Deployment and analysis of lecture recording in engineering education', *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*, Madrid, pp. 1–5. https://doi.org/10.1109/fie.2014.7044281

Witthaus, G. R. & Robinson, C. L. (2015) *Lecture capture literature review: A review of the literature from 2012–2015*, Centre for Academic Practice, Loughborough University, Loughborough. Available at: https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/25712/3/Witthaus_Lecture

Appendix 1

Calculation of adjusted minutes viewed

The following four values were first calculated:

- The average percentage viewed, averaged over all recordings and all students where the students only viewed and did not download (ClassV)
- The average percentage viewed, averaged over all recordings and all students where the students both viewed and downloaded (ClassVDL)
- The average percentage viewed by an individual student based on all recordings where the student only viewed and did not download (StudentV)
- The average percentage viewed by an individual student based on all recordings where the student both viewed and downloaded (StudentVDL)

The value of adjusted minutes viewed was then calculated as follows:

- For recordings that were viewed only, the adjusted minutes viewed equals the raw minutes viewed.
- For recordings where both access methods were used by an individual student, the adjusted minutes viewed for that student equals the raw minutes viewed times their view-only average divided by their viewed and downloaded (V&DL) average; that is, the number of minutes is adjusted pro rata based on their individual viewing data. If no view-only average is available for the student, the class values are used instead.

Adjusted minutes viewed = Raw minutes viewed * (StudentV/StudentVDL) or
Adjusted minutes viewed = Raw minutes viewed * (ClassV/ClassVDL)

- For recordings that were only downloaded by an individual student (i.e. not also viewed in the browser), the adjusted minutes viewed for that student equals the recording duration multiplied by their student view-only average, if available, or else the view-only average for all.

Adjusted minutes viewed = Recording duration * StudentV or
Adjusted minutes viewed = Recording duration * ClassV

Definition of the learning period and the revision period

The period between the date of each lecture and the examination is divided into two equal periods, designated the learning period and the revision period. The initial access by each student of each LC recording is placed into the appropriate period (learning or revision), and a 'number of days' value is calculated. For accesses in the learning period, this is the number of days after the lecture took place; for accesses in the revision period, this is the number of days after the start of the revision period. This methodology was required because the length of time between the end of teaching and the examination varies greatly between modules at Imperial College.