Student Engagement: An evaluation of the effectiveness of explicit and implicit Learning Analytics

Ed de Quincey University of Greenwich, Ray Stoneham University of Greenwich

Introduction

Retention and the measurement of student engagement are long standing problems within HE (Jones, 2008). A number of studies have investigated how students at risk of failing or withdrawing from University courses can be identified (e.g. Rugg et al.) but the issue still remains, and looks set to be a key concern with the rapid development of MOOCs within the sector (Yuan and Powel, 2013). One area of research in this field that is receiving increased interest is the use of implicit data collection and analysis, commonly known as Learning Analytics (LA).

LA has been defined as a method for "deciphering trends and patterns from educational big data … to further the advancement of a personalized, supportive system of higher education." (Johnson et al., 2013). Traditionally a student's progress and level of engagement has been measured by assessment, an explicit measure reliant on a student's submission of work and an academic awarding a mark and/or giving feedback, usually at the end of a module/course. However, in a student's day-to-day interactions with a University, other real-time implicit measures are being generated that are currently not being fully utilised e.g. attendance, VLE interaction data, library usage data, Web 2.0/social media usage. These measures do not normally require direct user interaction to record their engagement and so have advantages over labour intensive measures, such as exams and coursework.

HE already gathers an "astonishing array of data about its 'customers'" but has traditionally been inefficient in its data use (Siemens & Long, 2011). The analysis of this data though has the potential to identify at-risk learners and provide intervention to assist learners in achieving success (Macfadyen & Dawson, 2010) and so increasingly, student data is being aggregated and presented to tutors in the form of a Dashboard e.g. the University of Southampton's "Student Dashboard" (JISC, 2011). However, the representations being used are often based upon the ability of the developer to extract information from disparate sources and not on the types of data and interpretive needs of the user, usually falling far short of their potential (Few, 2006).

Making information available and transparent to tutors is only the first step however. Presenting student data back to students, using student centric formats and metaphors could tackle students' inability to access a composite, over arching view of their current learning activity, which can impact on a student's ability to develop creative divergent thinking skills (Rugg & Gerrard, 2009). A related issue that is frequently reported is students' inability to link skills that they are being taught on different courses together and how that impacts on both their employability and financial outlook.

A number of projects have investigated the use of LA and information representation/visualisation such as the Open University's Anywhere app which includes a "range of analytics that show how students engage with it" (theguardian.com, 2014), the University of Bedfordshire's student engagement system (n.a., 2012) and London South Bank University's partnership with IBM to "use predictive analytics to gauge if they might be falling behind" (Perry, 2014). However, there have been few studies that have systematically identified sources of pre-existing data and metrics from systems that are currently in use within a HE setting.

This paper explores the potential sources of data that represent/determine a student's level of engagement and progress by analysing the usage data of a bespoke Managed Learning Environment (MLE); the School of Computing and Mathematical Sciences (CMS) Intranet (Stoneham, 2012).

Methodology

The CMS Intranet has been incrementally developed since 2002 and contains the key information that a student needs in order to complete their courses. This includes the coursework specifications as Word or PDF documents, previous exam papers, screencasts and podcasts of some lectures, book lists, common teaching material, final year project documentation and relevant forms such as those for

extenuating circumstances, ethical approval and general enquiries. Very few handouts are given to students so learning materials are only accessible to them once they are logged in to the intranet.

Assessment is also supported on the intranet by student submission pages that record all coursework submissions and hand-ins. Students can submit work as many times as they like and the last one before the deadline is marked (with automatic penalties for late submission). Lecturers are able to provide feedback to students in a number of ways including annotated versions of the student's upload, screencasts with visual and audio feedback, recordings of project pitches and demonstrations as well as general textual feedback and provisional grades. The CMS Intranet also includes online eSupervision and Personal Tutoring Systems where most interactions between a student and their project supervisor and personal tutor are recorded. This includes meetings, messages, uploads of draft documents and reflective blogs.

All student interaction with the CMS intranet is recorded in the form of server logs. When a user requests a file from a web server, an entry is recorded in a log file i.e. by loading a web page via a web browser, a user is making a request for a HTML file along with other files that are embedded components of that page such as images and videos; each of these file requests make an entry in a log. These server log entries contain information such as the name of the file that was requested, the address of the page that referred the user to the requested page, the IP address of the device that requested the file (this can indicate the location of the user) and the time the file was requested. Tools such as Google Analytics are commonly used to analyse user traffic on a website but these require the embedding of a piece of code within a web page, something that cannot be put into a file such as a Word document. This means that non-web page file views and downloads cannot be tracked, which is where the majority of interactions are being generated on a VLE such as the one described above e.g. students viewing lecture slides in the form of PowerPoint files (.ppt), listening to audio feedback via MP3 files (.mp3).

As part of this study, functionality has been developed that takes this server log information and inserts it into a database, facilitating easier querying and analysis. Each server log record is simplified to contain the following information; the date and time of the request, the page/file requested and the userID of the user that requested the file. This userID is the email ID of the user supplied by the University and can therefore be linked directly with a student and their related data such as attendance, their course and modules of study, their tutor and lab groups and their assessment details. The combination of this data has then been used to compare traditional metrics such as attendance and coursework marks with interactions with the CMS Intranet. This has enabled an evaluation of the value of these measures in determining a student's progress and whether they can be used to identify "at risk" students, at various points in the academic year. The results of this comparison are described in the following section along with an overview of the general usage of the intranet.

Results

Server log data generated by 3,576 students across the School since September 2011 to June 2013 has been collected and during this time there have been 7,899,231 interactions with the CMS student intranet. For this study, the period from September 3rd 2012 to May 29th 2013 has been analysed, to represent an academic year, with 2,544,374 interactions from 2,634 students being recorded.

The distribution of interactions over the academic year is shown in Figure 1 below, showing that there are peak times during the year where the intranet is being used.
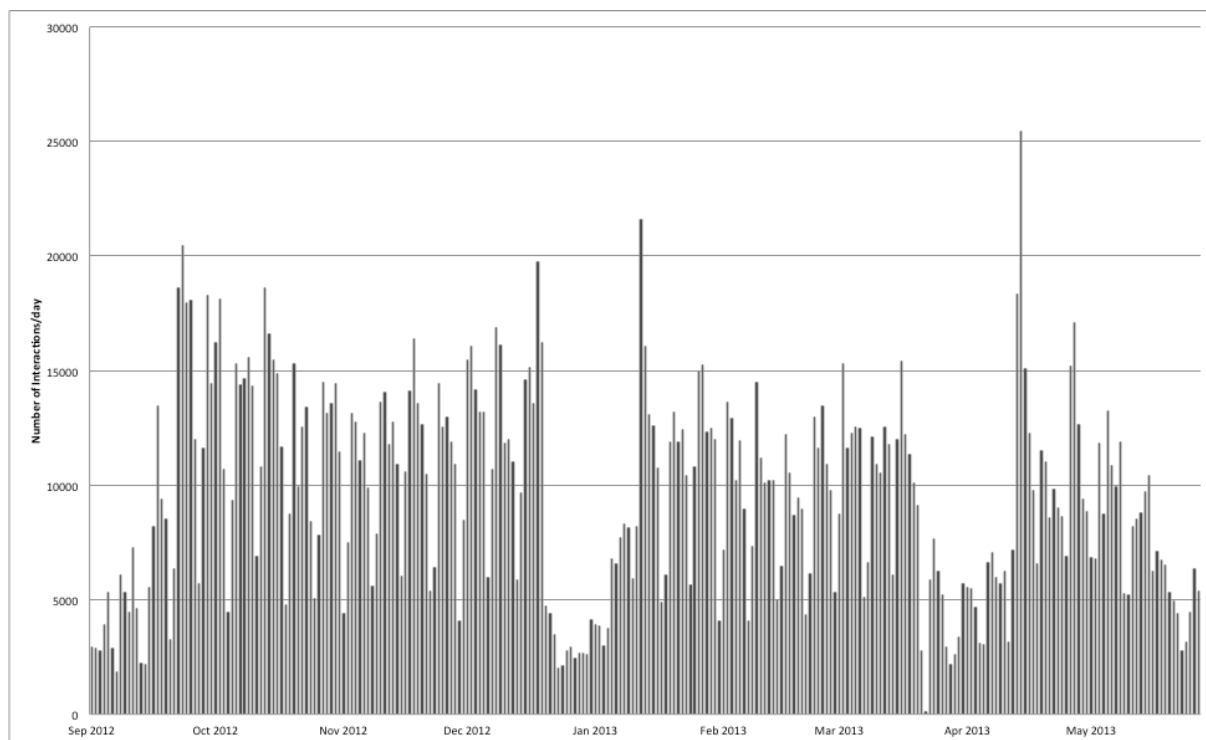
Figure 1: Distribution of activity on the Intranet per day during the Academic year 2012 to 2013

These relate to the start of each semester at the end of September and middle of January and also during coursework deadlines towards the end of a semester. There is also generally less activity on weekends than in the week and reduced activity during holiday periods.

File Types

The following table shows the number of interactions/downloads for the most popular/relevant file types.

File Type	Web page	Word	PDF	PowerPoint	ZIP	MPEG	Excel	MP3
Number of Interactions	2,131,278	157,607	128,676	66,129	19,561	18,974	6,368	1,851

Table 1: Number of interactions with different file types on the CMS Intranet

The majority of views were of the web pages of the intranet itself, representing the large number of tasks that the site supports. These can then be analysed further in relation to the different resources on the intranet that students are (or are not) using. For example, there were 18,507 interactions with the section of the intranet that provides access to past exam papers by 911 individual students (35% of the total number of students)[1] and 3,181 views of the page that contains advice on plagiarism by 778 students (30%). This information is useful for the staff responsible for these areas in determining whether the content is being utilised by students and also is an opportunity to target those students who are not using these resources with reminders/instructions about how to access these materials. This point is discussed further in the discussion section.

Of the other file types, these are predominantly resources created by staff to support teaching, learning and assessment. The majority of the views of Word documents (4,539 individual files) are coursework specifications (15%) with one particular specification receiving over 2,500 views by 243 students. Interestingly 20 registered students did not look at the coursework specification for this module and 103 students (39%) viewed the specification more than 10 times (this is explored further in the following section). The views of PDF documents were also comprised of coursework specifications but the most downloaded file was an iBook created to support students on a Level 2 Digital Media Design course (1,390 views) followed by a number of example final year projects that had been

---

[1] This low percentage can be explained by some modules not being assessed by exam.

uploaded to the project resources area of the intranet (5,883 views of 13 project reports by 430 students).

Comparison of Measures

In order to identify which implicit measures might determine/represent a student's progress, two Computing Undergraduate modules have been considered; a first year module called "COMP1314: Digital Media, Computing and Programming" and a third year module "COMP1640: Enterprise Web Software Development". COMP1314 is a 30 credit introductory course to computers and programming, assessed by 2 pieces of coursework and an exam. COMP1640 is a 15 credit final year course that aims to give students practical experience of developing enterprise systems using web technologies and is assessed by a piece of group coursework. Both modules contain students with similar levels of ability but at the opposite ends of their degree experience.

For both of these modules, comparisons between the student attendance, final mark and intranet activity, categorized into various resource types, have been made to investigate which factors might affect student engagement and progression.

COMP1314 (first year module): Comparison of marks, attendance and intranet activity

For this module, Figure 2 shows strong positive correlation was found between the final module mark and overall attendance at tutorial and lab sessions (0.64).
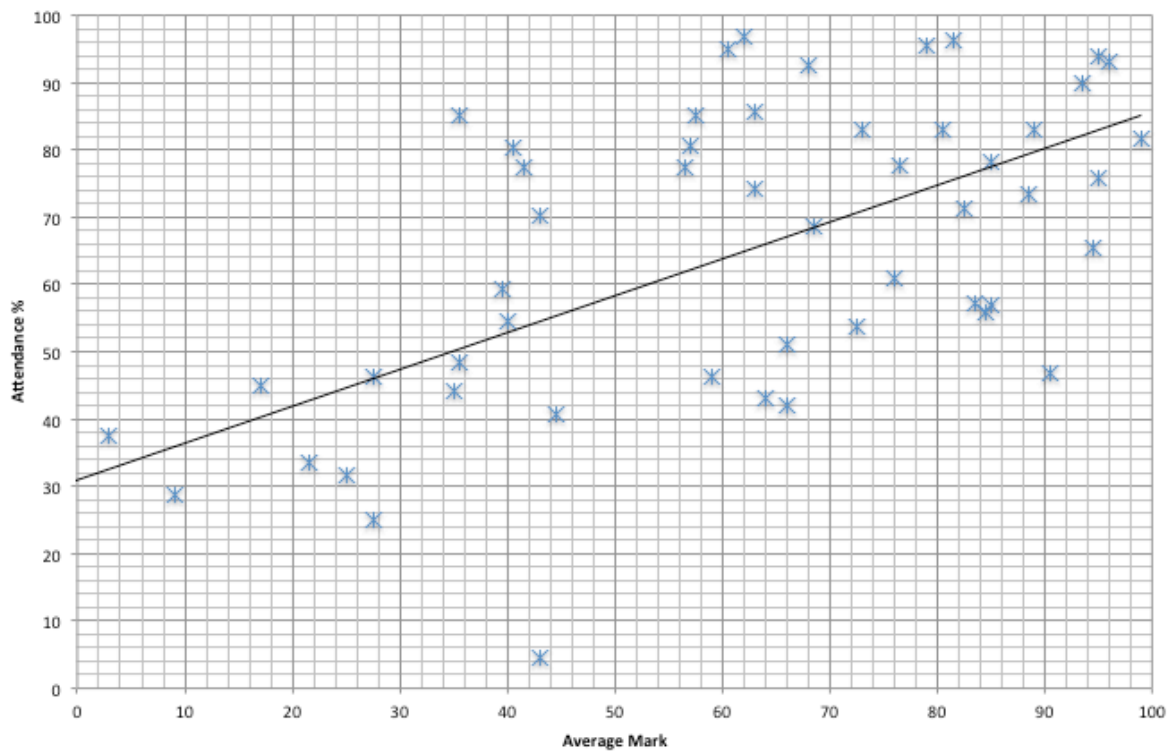

Figure 2: Relationship between a student's average final mark and their attendance percentage (n=53, Correl. = 0.638)

Figure 3 below shows equally strong positive correlation between the final module mark with the number of total intranet interactions during the year (0.6). This correlation increases slightly (to 0.63) when only intranet resources related to COMP1314 e.g. views of lecture slides, tutorials, coursework specifications, reading lists etc. are considered. On average, a student interacted with COMP1314 intranet resources 215 times.
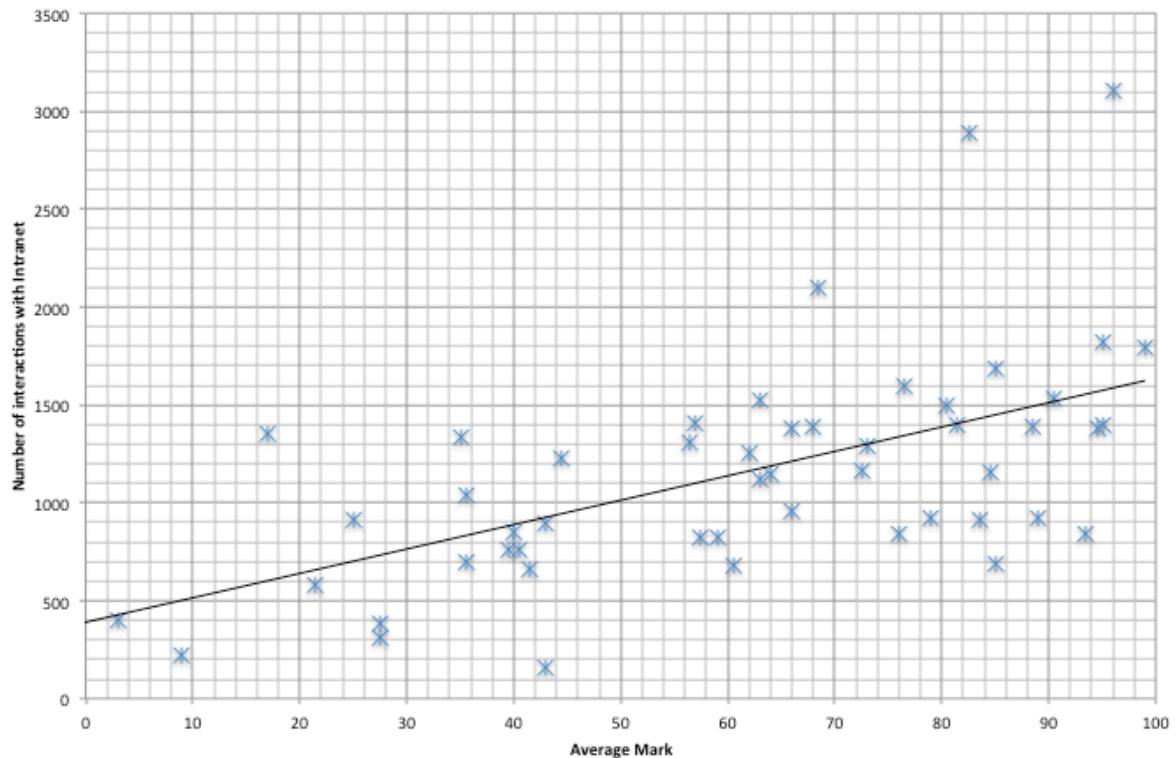
Figure 3: Relationship between a student's average final mark and their intranet activity (n=53, Correl. = 0.601)

This indicates that students that have high levels of activity both physically and virtually with the module tend to have higher marks and students with low levels of activity, achieving lower marks. When considering views of lecture slides and tutorial instructions separately, each lecture slide handout has been viewed on average 76 times whereas each tutorial instruction has been viewed on average 142 times.

There is also strong positive correlation between the number of intranet interactions and a student's overall attendance (0.44), perhaps countering the generally held belief that making materials/services available online decreases attendance in lectures.

Interestingly there was a weak positive relationship (0.23) between the number of times the coursework specification had been viewed and a student's final mark, with the specification being downloaded by each student 7.4 times on average.

COMP1640 (third year module): Comparison of marks, attendance and intranet activity

For this module, there was similar, moderate positive correlation between attendance and the final mark (0.42) but as shown in Figure 4 below, weak/negligible correlation between the interaction with module resources/pages and final mark (0.18). On average, a student interacted with COMP1640 intranet resources 119 times.
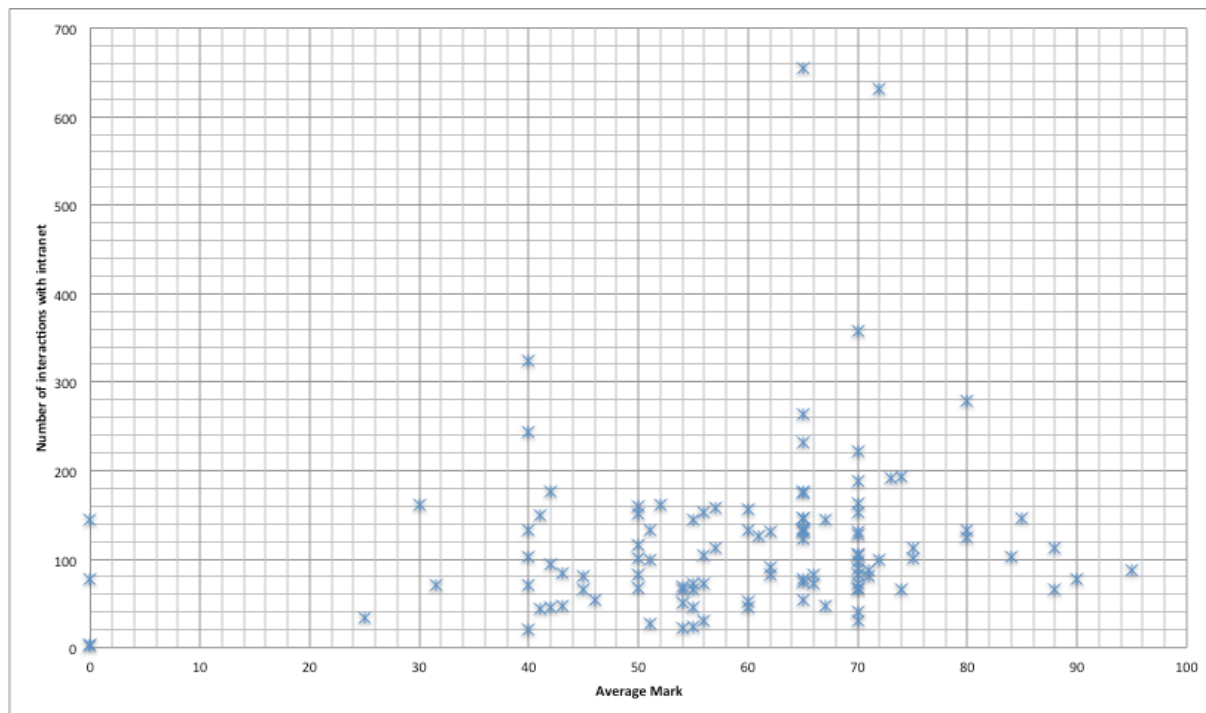
Figure 4: Relationship between a student's average final mark and their intranet activity (n=109, Correl. = 0.18)

There was in fact no relationship between views of module lecture materials and the final mark (-0.07). Whether this reflects improved digital literacy i.e. saving instead of downloading the same file multiple times, less reliance on module materials or simply the nature of the module is currently being investigated. However, there was a moderate positive correlation (0.38) between the number of times the coursework specification had been viewed and a student's final mark and on average each student downloaded the coursework specification 9.4 times.

Firsts and Fails: Temporal distribution of intranet activity

As shown in Figure 5, the distribution of intranet activity shows that the pattern of usage is similar to begin with for students on COMP1314 that eventually receive first class marks and those that fail, with relatively high levels of activity during October and November and a decrease in December[2]. First class students then have a similar patter of activity to that in the first semester whereas failing students tend to remain at low levels. On average, failing students have half the number of interactions with the intranet than first class students throughout the year.

---

[2] This has been calculated by taking the average number of interactions at different points of the year for groups of students that are awarded first class and failing marks.
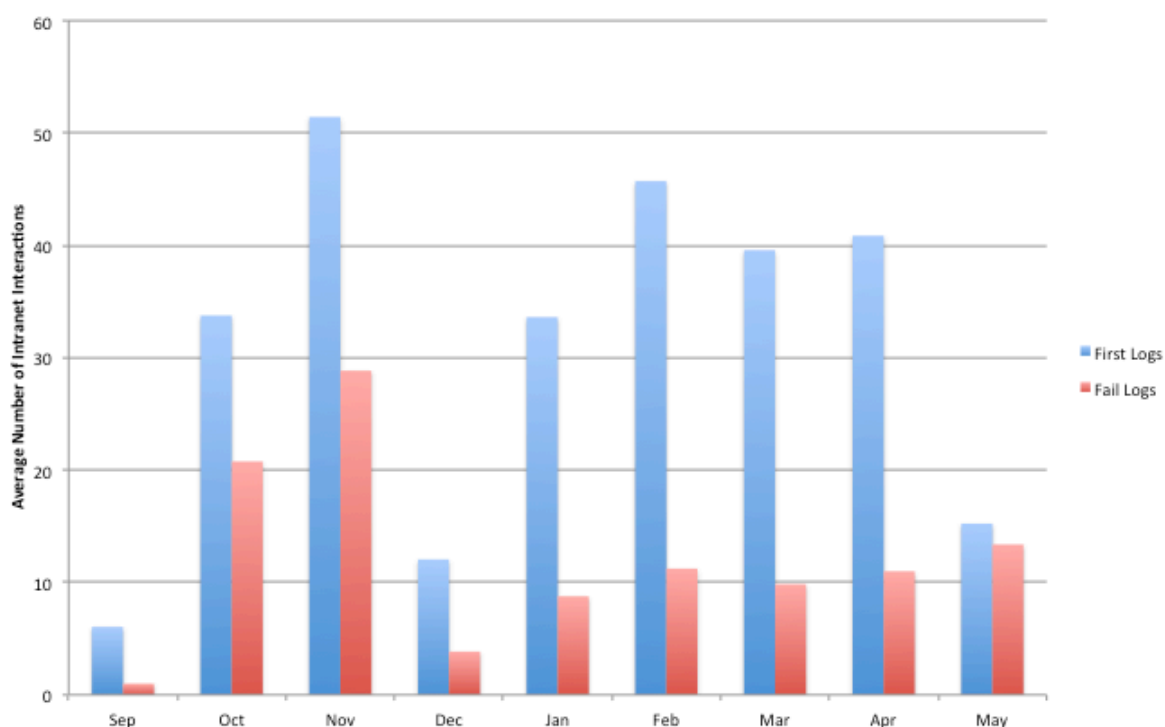
Figure 5: Distribution of intranet interactions for and average "First" student and an average "fail" student

Table 2 below shows the averaged profile of a student who is awarded a first class mark on COMP1314, compared to an averaged profile of a student who fails the module. For each set of activities, both physical i.e. attendance and virtual i.e. intranet interactions, a first class student has double the amount of activity as a student that fails this module.

	Average Mark	Attendance	Total Intranet Interactions	Intranet Files Downloaded	COMP1314 Interactions	COMP1314 Lec./Tut. Views	CW Spec. Views
Average "First" Student	86%	75%	5.3 per day	213	278	103	8
Average "Fail" Student	21%	40%	2.6 per day	121	118	52	5

Table 2: Average profile of students who are awarded First class marks (>=70%) and Fails (<40%)

Discussion

The results presented in the previous section have shown that there are clear differences in the levels and types of online activity between students and that these levels may be factors in determining the final levels of achievement within modules. The following sections discuss potential explanations for these findings and suggest possible uses and implication for LA and the integration of these measures in MLE's and predicative/automated systems.

Attendance

A number of previous studies have demonstrated the importance of attendance and the affect this has on final grades e.g. (Schmidt, 1983; Park & Kerr, 1990; Ryan et al., 2010). Although this study has not taken into account other important considerations such as pre-entry qualifications, attendance on the first year module, COMP1314, showed a positive correlation with a student's final grade at similar levels to previously reported accounts for first year science degree courses e.g. Gunn (1993) reported a

correlation of 0.66, with similar levels seen in the final year course. The implication is that for any LA based system, physical attendance is an important metric for measuring engagement and a predictor of future achievement. Further work is currently being undertaken with the full data set of all students recorded in the CMS Intranet to determine whether the temporal distribution of attendance is a significant factor and at what point in the academic year future achievement can be predicted.

Intranet Activity

The differing levels of correlation found between the two modules and the types of resources that were viewed show that further consideration needs to be made for which measures are suitable for LA and to what extent they impact on a student's final grade. For the first year module, similar levels of correlation to that of physical attendance were found for students viewing module materials and their final mark in that module (in this case lecture slides, audio recordings of some lectures and tutorial instructions). At this level, it seems that viewing module related materials online is as important as lecture and tutorial attendance. The fact that 25% of the assessment of this course is based on weekly tutorial uploads, where the tutorial instructions are only available via the CMS Intranet perhaps explains the fact that they were viewed on average twice as much as lecture handouts and further analysis into comparing the marks for that component of the module and views of tutorial instructions is currently being undertaken. However, beings as a student's progression is largely determined on them completing these weekly exercises, this is clearly a metric that should be considered when designing a LA system i.e. which files/resources are essential for completing the module assessment should perhaps receive a greater weighting when building predictive algorithms, with students and lecturers receiving alerts if these files are not being viewed.

For the final year module, COMP1640, it was interesting that no correlation was found between module lecture material views and a student's final mark. The average attendance percentage of a student on this course of around 65% is comparable with that for COMP1314 of 63% and the average number of interactions with module pages/materials is similar (taking into account the fact that COMP314 runs for two semesters). The main difference between the two groups therefore is other material that was made available to students with the majority of views on COMP1314 relating to the weekly tutorials that were part of the assessment. For COMP1640 there were no equivalent weekly tasks apart from suggested weekly meetings with a tutor and the group of students that they were working on the coursework with (all details of which were in the very detailed specification). This may explain the fact that there was a moderate correlation between the number of times the coursework specification was viewed and the final grade and the fact that it was viewed twice more on average per student than COMP1314, in half the amount of time.

For both of these courses, it is clear though that there is a link between views of the files relating to assessment and a student's final grade, which has implications for the metrics used for LA and perhaps shows that online activity on an MLE is more centered around assessment.

The general behavior observed of students in both groups downloading files repeatedly instead of saving them to their own areas/drives or printing them e.g. the coursework specifications, perhaps indicates either lower levels of expected digital literacy or a shift in students' perceptions of where files are stored and viewed, due to the ubiquitous nature of cloud based services e.g. a user will tend to take a photo with an application such as Instagram and upload it and view it online as opposed to saving it and copying to multiple devices.

Profiles

A potentially surprising finding form this study is the distribution and amount of intranet activity that first class students display compared to students that fail. This study has shown that failing students still engage with a module but at around half the level of a first class student. Over an academic year though, these levels tend to fall off more notably in the second semester until May, where activity recovers to similar levels (this can perhaps be explained by students looking at materials for the exam and then resit courseworks). As far as implication for LA are concerned, this immediate difference in online activity at the start of the academic year should be considered and appropriate interventions planned e.g. checking to see whether students know where the materials are or allowing them to view their own usage data in comparison to students at different predicted levels.

From the perspective of a module coordinator or personal tutor, being able to view how resources are, or are not being used in real time has substantial implications for module delivery and pastoral care. Beings as attendance at lectures and tutorials is a commonly accepted, highly visible indicator of engagement, being able to see a student or cohorts' level of engagement with the non visible components of a module such as lecture handouts, tutorial materials and reading lists, could be a significant factor when judging a student's profile or the success of the module's delivery.

Future Work

Further analysis of this data is ongoing and a full evaluation of all 3,576 students is planned using techniques such as Bayesian Belief Network Analysis. Specific areas of focus however include an examination of when and how often students access their feedback (either text based, audio or video) relative to release dates, and whether this is a key indicator of their engagement and desire to benefit from feedback. Also the record of interaction between a student and their supervisor and personal tutor via the eSupervisor and Personal Tutor Systems might indicate whether a student has been taking advantage of the support that the school offers in these areas and will therefore be investigated. Finally, preliminary results show that only around 30% of students view pages and information related to plagiarism, that are on the CMS Intranet. Whether this is because students are unaware that this information is being made available or they do not see its importance will be explored along with methods for highlighting it to students.

Conclusions

The results from this study indicate that attendance and interactions with a student intranet are useful measures for student engagement and predictors of success, particularly in a student's first year. Reasons for the difference in effect observed between first and third year modules have been tentatively identified, and further investigation is currently being undertaken on the full data set. This work shows that there are clear implications for LA, and for educators in general, regarding expected patterns and levels of activity for different types and levels of student and that increased emphasis may need to be placed on measuring interactions with assessment based materials.

References

Few, S. (2006) *Information Dashboard Design: The Effective Visual Communication of Data*, O'Reilly Media, Inc., Sebastopol

Gunn, K. P. (1993) *A correlation between attendance and grades in a first-year psychology class*. Canadian Psychology, 34, 201–202

JISC (2011) *Southampton Student Dashboard*. [online] Available at: < http://www.jisc.ac.uk/whatwedo/programmes/bce/relationshipmanagement2/dashboard.aspx > [Accessed 15 April 2013].

Johnson, L., Adams Becker, S., Cummins,M., Estrada,V., Freeman,A., and Ludgate, H. (2013) *NMC Horizon Report: 2013 Higher Education Edition*, Austin, Texas:TheNew MediaConsortium.

Jones, R. (2008) *Student retention and success: a synthesis of research*, York: HEA

Macfadyen, L.P. & Dawson, S. (2010) *Mining LMS Data to Develop an 'Early Warning System' for Educators: A Proof of Concept*, Computers & Education, vol. 54, no. 2, pp. 588–599.

n.a. (2012) *Supporting institutional decision making with an intelligent student engagement tracking system*. [online] Available at: < https://jiscinfonetcasestudies.pbworks.com/w/page/60129349/University%20of%20Bedfordshire%20-%20Business%20Intelligence > [Accessed 9th June 2014].

Park, K.H. & Kerr, P. (1990) *Determinants of Academic Performance: A Multinomial Logit Approach*. Journal of Economic Education 21: 101-111.

Perry, K. (2014) *New IT system will identify failing university students*. [online] Available at: < http://www.telegraph.co.uk/education/educationnews/10805717/New-IT-system-will-identify-failing-university-students.html > [Accessed 9th June 2014].

Rugg, G., D'Cruz, B., Foreman- Peck, L., Grimshaw, E., Guilford, S., Roberts, D., Tonglet, M.,

(2008) *Selection and Use of Elicitation Techniques for Education Research*, International Journal of Information and Operations Management Education, 2 (3), pp.235-254, ISSN 1744-2303.

Rugg, G. & Gerrard, S. (2009) *Choosing appropriate teaching and training techniques*, International Journal of Information and Operations Management Education, 3(1), pp. 1- 11.

Ryan, M., Delaney, L., & Harmon, C. (2010). *Does lecture attendance matter for grades?: Evidence from longitudinal tracking of Irish students*. Belfield: University College Dublin.

Schmidt, R.M. (1983) *Who maximises what? A study of student time allocation*, The American Economic Review, 73:23-28.

Siemens, G. & Long, P. (2011) *Penetrating the Fog: Analytics in Learning and Education*, Educause Review 46 (5), 30-32

Stoneham, R. (2012) *Managed Learning Environments in Universities: Are they Achievable?*, Compass (the Teaching and Learning Journal of the University of Greenwich), vol. 6, pp. 45-54.

theguardian.com (2014) *Winner: The Open University Project: OU Anywhere app.* [online] Available at: < http://www.theguardian.com/higher-education-network/2014/feb/18/winner-the-open-university-distance-or-online-learning > [Accessed 9th June 2014].

Yuan, L., Powell, S. (2013) *MOOCs and Open Education: Implications for Higher Education,* JISC CETIS

Authors

Dr Ed de Quincey
University of Greenwich, Department of Computing and Information Systems
Old Royal Naval College, Park Row, London SE10 9LS UK
e.de.quincey@gre.ac.uk
Dr Ray Stoneham
University of Greenwich, Department of Computing and Information Systems
Old Royal Naval College, Park Row, London SE10 9LS UK
R.J.Stoneham@greenwich.ac.uk