Tracking our students as they learn

Eric Bodger Eric.Bodger@winchester.ac.uk
Mike Hart Mike.Hart@winchester.ac.uk
David Rush David.Rush@winchester.ac.uk

Business Management Group
Faculty of Social Sciences
The University of Winchester
Winchester
SO22 4NR
Tel: +44 (0)1962 827 323

Abstract

The advent of e-learning provides us with a range of new ways to track the learning activities of our students, alongside well-established ways from the past. Here we explore how some simple tracking mechanisms can be used to predict a student's eventual performance. The intention is thereby to provide for the student and tutors early indicators of poor performance and withdrawal from the course. Previous investigations have concentrated upon attendance as an indicator of performance. We have repeated this approach in a somewhat different subject than was used for the earlier investigations and have also found a strong correlation between the two. Another possible performance indicator may be derived from interactions with a Virtual Learning Environment. Most studies of such interaction have looked at these in detail to gain understanding of the learning processes in online forums. For the purposes of performance prediction we have chosen a simple measure, the total number of interactions, and have found this also to be correlated with performance. This raises the possibility that a simple measure of activity that is readily obtainable from most VLEs might be used in managing courses to achieve overall better student performance and higher retention rates.

Keywords: tracking, attendance, VLE interactions, performance

1. Introduction

The advent of pervasive technology allows us to track the details of people's movements and actions through their use of mobile phones, Internet access, CCTV footage and so on. Students have always been subject to some form of tracking during their learning. In pre-electronic days it might have been simply a record of attendance, augmented perhaps by the results of formative tests. With the advent of e-learning there are considerable possibilities for the details of student interactions with their tutors and other students through Virtual Learning Environments (VLEs), emails, forums, blogs and Personal Response Systems (PRS) to be stored for retrieval and analysis. This paper will explore the utility of such tracking mechanisms, through a survey of reported experience and presentation of the results of investigation of this issue within our own institution.

There have been many investigations using tracking data for a variety of purposes (Colby 2004, Kirkpatrick 2005, Paananen and Simpson 2007, Wang 2004). Some have analysed the detail of student electronic interactions in forums in terms of models of the learning process. Others have used the data to compare the delivery and outcomes between traditional and electronic delivery of the same course. Our interest is somewhat different. We are interested in whether simple measures derived from tracking data can help to support ways of increasing student engagement. We believe that increasing engagement will be reflected in improved performance. In this paper we take two items as signifying engagement. Because we are dealing with blended learning courses, one of these is physical attendance. The other is the total number of

interactions with a VLE. We are interested to know whether in a variety of courses there is a link between engagement defined in this way and student performance. If such a link can be established, it would point towards approaches that could be taken to use tracking data to improve course delivery. This would include acting as diagnostic information to be used in supporting students, and helping tutors in reflecting on their own practice. It will also be noted that there is a possibility that such tracking will alter, for better or worse, the personal relationships between students and tutors.

2. The Uses of Tracking

2.1 Tracking Attendance

There have been several studies of a possible link between attendance and performance. In part they have been undertaken due to a concern of academic staff about the ready availability of teaching material on VLEs in a blended learning environment. It has been suggested that this will lower students' attendance at teaching sessions and hence their overall performance. Here we consider three such studies undertaken in UK universities. They all explored this link for undergraduate students on technical courses. The first is that of Colby (2004) who investigated the relationship in a first year programming module of over 100 students. His graphical and tabular results show a very clear correlation between attendance and performance. An interesting related result from this study was that there was a steady fall in the average attendance during the semester. Colby suggests that alarm points for action can be deduced from these results and could be presented in the form of rules to students to make them aware of the consequences of absence. One such rule is the Seventy Percent Rule – if a student does not attend at least seventy percent of teaching sessions they have a two in three chance of failing, and a four in five chance of not getting a first or upper second. From the viewpoint of the tutor the Week Two Rule is important. This spells out a level of non-attendance, for whatever reason, during the first two weeks that is a cause for concern. Accepting the significance of this would imply that attendance data from those first two weeks could be used to drive remedial action.

The study by Burd and Hodgson (2005) followed the general approach of Colby, but on a larger scale. They examined attendance on five second year modules of a Computer Science degree over a five year period with a cohort size of 70. Their findings supported those of Colby. There was a significant correlation between lecture attendance and examination performance for all five years. A comparison of the attendance of those students gaining a module mark above 60% with those below this level showed the former to generally have attended a greater proportion of their lectures. Again the results supported Colby's Seventy Percent and Two Week Rules. In considering this agreement between the two studies it should be noted that the student ability profile and the organisation of the teaching were markedly different in the two institutions.

The paper by Barrett et. al. (2007) also investigated attendance in a year one programming module. In contrast to the two previous studies no correlation was found between attendance and performance. This was attributed in part to the recorded attendance relating to tutorials and practicals only and the assessment mark being for the first three assessments out of five for the module. The result was discussed in the context of replies to questionnaires from both staff and students. These showed that for this group of students their approach to the module was affected by factors such as the need to have employment. The impact of having study notes readily available on a VLE was explored. The majority of tutors felt that doing this was adversely affecting students' attendance but only a small proportion of students felt that their attendance in lectures and tutorials was influenced in this way.

.

These three studies were based on students studying Computing degrees. This raises the question of the applicability of these findings to other subject disciplines. It might be argued that the topics studied on a Computing degree such as programming are

highly linear, that to miss the content of one week upon which subsequent development is based would increase the risk of poor performance. In that case the effect might not be so marked in less technical subjects. This is an issue that could benefit from further investigation.

Another factor that could affect performance is student gender – see for example Woodfield et. al. 2005. What does not seem to have been investigated is whether gender also affects attendance, and how these three factors are jointly related. We consider some results on this in section 3 below.

2.2 Tracking VLE Interactions

There have been several attempts to produce a model of the interactions that take place in an online learning environment and to use such models in the analysis of tracking data. One of these is the 5 factor model of Pozzi et al. (2007). The model, a development of earlier work, proposes analysing online discussions in terms of 5 'dimensions': participative, interactive, social, cognitive and meta-cognitive, and teaching. The utility of such a complex model is not obvious. It is assumed that the designation of a single interaction to one of the 'dimensions' will be done by content analysis. Yet the definitions of the 'dimensions' would not seem to allow this to be done unambiguously. For example the participative dimension includes 'passive participation', whilst the interactive dimension includes 'passive participation before posting'. More insight into what is implied by the use of content analysis is provided by the paper of Lally (2002). The learning processes of a small group of only eight postgraduate professional students on a distance learning module were tracked through their contributions to an online discussion forum. The analysis of their interactions was through coding schemes assisted by data analysis software. In fact results for two different coding schemas were produced, one for the learning type of messages and another for the tutoring type of the messages. However, even for this small number of students the coding was a significant task and coding was restricted to a sample of 10% of the messages spread equally between the beginning, middle and end of the activity. This restriction was in part due to the complexity of the adopted coding scheme which used a total of 60 categories and sub-categories. There were also cases in which messages could not unambiguously be attributed to a single category, requiring discussion between the researchers.

Of somewhat more relevance to our work are those studies where the results of delivery by blended learning and e-learning are compared using more traditional methods. An example is that of Paanen and Simpson (2007) who compared the performance of a small group of postgraduate Accounting students tested after sections of their module had been delivered by traditional mode and then by a flexible delivery system (SMIRK) allowing presentations and linked audio to be delivered over the Internet. They found significant improvement on the tests of that part of the module delivered by SMIRK. Another result from this study was that use of the VLE, available throughout the module, increased significantly when tuition was through SMIRK.

However the small numbers involved and the fact that the two sections covered different material (possibly of different difficulty) mean that this result must be treated with caution. Like many of the studies comparing traditional and blended or e-learning this one has confounding variables that may have influenced the results. In this case these are clearly acknowledged by the authors but this is not always the case.

Another study of a small group of postgraduate students is that of Wang (2004). The module was completely online. The analysis was in terms of a 'Cybergogy for Engaged Learning' – a model that stresses the need "to encourage students' cognitive, emotive and social presence in an online learning environment". Of direct relevance to the results reported later in this paper is the finding that total usage of the VLE was highly correlated with performance as measured by module grades. Apart from total usage this paper also used a measure entitled 'visibility score' for students, constructed from

the number of their interactions such as posts to forums, contributions to chats and so on. This measure was also correlated with performance.

Here we are exploring the use of tracking for somewhat simpler purposes. Our focus is on the potential of tracking data for predicting student performance and guiding interventions by tutors in supporting students. We are not looking at using tracking data to understand, or to assess or to model the details of the activities that a student undertakes; rather we are interested in establishing what broad factors affect the performance of a student on a task such as the production of an assignment or in an exam.

3. Tracking in Practice

The objective of this investigation was to explore whether readily obtainable measures from tracking data are related to student performance. If this is the case then it should be possible to use those measures to aid students, indicating to them whether their activity means they may be in some 'at risk' category, and helping them to adopt ways to improve. So our intention was to address course management issues, rather than to investigate student learning processes. In fact, for some time our Group has tracked student attendance, and taken action in the belief that poor attendance is correlated with poor performance. An empirical justification of this supposed relationship in the context of our particular students was another intended outcome of this research. In a blended learning environment many different measures derived from tracking data might be used because electronic systems record a wealth of data about interactions between the student and the system and with other students and tutors. With the objective of using tracking measures that are easy to obtain we concentrated on just two in this work, the simplest we could envisage. If these are found to be sufficient there will be no need to explore more complex measures for the purposes we have outlined here. For our investigation we had available data for the same cohort, measured in a Year One and a Year Two module – described in more detail below.

The simplest possible tracking data comes from a student's physical attendance at a teaching session. Obviously this is not a measure that relates only to a blended learning environment but it can be readily collected in both blended and traditional settings. The collection and analysis of this data does however require resources and in a large class data collection is usually less than 100% for identifiable reasons (Colby 2004). It is conceivable that a hardware driven solution could be used, placing the onus on the student to register attendance. Possibilities include automatic fingerprint identification, but apart from the cost of the readers there are obvious objections to going down that path. Many tutors would see the introduction of tracking equipment of this kind as changing irrevocably the nature of teaching and learning in Higher Education. Whether or not that is true, tutors would certainly be less than enthusiastic and probably uncooperative about its use. In this study these problems of the accuracy of attendance data are not of great significance. This is because the teaching sessions took place in small groups (< 27 students) where attendance can be quickly and accurately checked. Thus we believe that the attendance data presented below has high integrity.

As will be seen the use that students made of a VLE was tracked in one of the two modules we investigated. Although students on the Year One module could download electronic material no great use was made of other VLE type facilities such as forums, so it was not thought relevant to this study to consider tracking data for that module. By contrast the Year Two module was delivered by means of Moodle (Dougiamas 2007), a well-known open source VLE with a good range of interactive facilities. In the delivery a number of these had been used including document repository, forums, chat, and student submission of assignments. It was not possible to take part in the module without at least a minimal use of the VLE since some classroom activities depended on it and assignments were submitted and returned via it. Moodle provides fairly comprehensive tracking data. For each activity a record is created giving time, name of

user, and type of activity, of which there are many. It would have been possible to base our investigation upon the data relating to one or a few of these types. Investigations such as those of Lally (2002) have concentrated on the nature of the contributions to forums. This was appropriate to the distance learning modules studied there, where the learning activity was closely identified with forum use. On our Year Two module, however, the forums were not the central focus of VLE use. They were there to provide for students to comment on and ask questions about the assignments and raise general questions about the module. Their only mandatory use was for some leaning activities in the classroom. But as other facilities than forums had been actively used and as the objective was to use simple measures, the number and nature of posts to forums was not used as the tracking data. Another reason for this was that posting to the forum is not the only way to get benefit from it. As Beaudoin (2002) has demonstrated, students who adopt a low profile approach to online learning and make a small number of posts to online forums, nevertheless feel that they are benefiting from their less active engagement. For these reasons the tracking data measure chosen for engagement with the VLE was simply the total number of interactions.

We now describe the two modules investigated here. They were both part of a Business Management degree, but had a technical bias. The Year One module was entitled Computer Applications, which introduced the main types of business decision support and communication packages. The total cohort was 91 students but this was taught in four separate groups. The Year Two module was an option entitled Business Decision Technology and had only 16 students. It concentrated on more advanced techniques used in Business to aid decisions. Both modules were therefore rather more technical than most Business modules, but were certainly not as technical as the Computing degree courses of section 2.1. It is worth noting that the 16 students on the Year Two module were all part of the Year One cohort of 91. Thus this study was based upon the same group of students at Year One and Year Two levels. Although the module delivery aimed for interactivity the underlying philosophy was not constructivist. There was no group working. The modules are typical of many on Business Studies degrees.

The measures used in this study were largely unobtrusive. Attendance records were obtained manually with the very minimum disturbance of classroom activity and there was no specific requirement to attend. At the time of collection there was no intention to use this data for research purposes, adding to the unobtrusiveness of the measures. Records of electronic interactions were kept automatically by the system.

3.1 Attendance

In this section we present the results of attendance data correlated against performance. The attendance data was collected as described above. The results for the Year One module are given in Table 1. This gives the correlation coefficient for the number of weeks attended (out of 12) against the overall performance on the module. The results for the whole cohort are followed by those for the four subgroups in which the students were taught (S denoting Single Honours and C Combined Honours).

The whole cohort results show a significant correlation and this is in line with the results of Colby (2004) and Burd and Hodgson (2005) which were also obtained with comparatively large group sizes (about 150 and 70 respectively). The whole cohort results are reflected in those for the sub-groups S1, S2 and C1. However in C2, admittedly a smaller group than the others, we do see a markedly different result. This may be a statistical artefact of the smaller numbers or there may be some other more subtle effects that would require more systematic investigation.

Table 1 Attendance versus performance for Year One module

Group	n	Pearson Coefficient	p-Value	Significant at 5%
Whole cohort	91	0.410	0.000	✓
Sub-group S1	24	0.467	0.022	✓
Sub-group S2	27	0.561	0.002	✓
Sub-group C1	25	0.468	0.018	✓
Sub-group C2	15	0.075	0.789	

The results for the Year Two module, correlating attendance and performance in both assignments and for the module mark (aggregated from the two assignment marks) are given in Table 2. There are significant correlations for one assignment mark and the module mark. It seems that even with the small numbers on this module there is again evidence to support the link between attendance and performance.

Table 2 Attendance versus performance for Year Two module

Assessment	n	Pearson Coefficient	p-Value	Significant at 5%
Assignment One	16	0.643	0.007	✓
Assignment Two	16	0.235	0.380	
Module Mark	16	0.531	0.034	✓

An analysis by gender of the Year One module data revealed some interesting results. The data was dichotomised into 'high attenders' (10 or more presences out of 12) and 'Low attenders' (9 or fewer attendances out of 12) The data was also dichotomised at the mark of 60 (Upper second and above, lower Second and below). These results are shown in table 3.

Table 3 Gender versus attendance versus performance for Year One module

		n	Marks	p-Value	Significant at 5%
Males	Low Attendance	28	56.5		
	High Attendance	22	60.5	0.22	
Females	Low Attendance	15	58.1		
	High Attendance	26	64.9	0.047	✓

Here it can be seen that although attendance does not show a marked association with marks in the case of *Males,* in the case of *Females,* there is a clear difference in the marks obtained by Females when classified by attendance.

3.2 Electronic interactions

In this section in Table 4 we present the results of the correlation of interactions with the VLE against performance in both assignments and for the module. For the reasons given above the total number of interactions with the VLE was used. This number varied considerably from student to student, with a mean of 345, but the total range was from 61 to 1585. The correlation coefficient is significant in all cases.

Table 4 Interactions versus performance for Year Two module

Assessment	n	Pearson Coefficient	p-Value	Significant at 5%
Assignment One	16	0.560	0.024	✓
Assignment Two	16	0.695	0.003	✓
Module Mark	16	0.674	0.004	✓

4. Conclusions

These results on the measures of attendance and number of interactions versus performance are admittedly limited at present. The results do however indicate that both these simple measures are powerful predictors of performance. They would therefore seem to have a role in the management of courses, particularly in relationship to identifying students whose lack of engagement with the course is likely to lead to poor performance.

At a practical level, a lot of time is spent checking attendance (engagement) of our students, especially at the start of a module. We therefore envisage that we will next compare attendance and online participation in the early weeks of a module to confirm the correlations suggested here, with the intention of using the results to achieve overall good cohort performance and retention rates. It seems from these results that the number of interactions in a VLE is as good an indicator of performance as attendance. Since it is much easier to get data about VLE interactions, confirming these results would help considerably in managing modules. It would no longer be necessary to collect attendance data.

.

We recognise that the numbers in some of these data sets are small, as indeed they are in some other cited investigations. We therefore consider these results to be suggestive rather than definitive. Further investigations will need to be based upon larger sample sizes.

As indicated above the data for this investigation has been obtained in an unobtrusive way. If the data were to be used as suggested to help students this might alter student behaviour. Students might interact with the VLE so as to create interactions and the personal relationships between students and staff might be affected, to the overall detriment of the course. This possibility would need to be kept in mind in future developments.

References

Barrett R, Rainer A and Marczyk O (2007) “Managed Learning Environments and an Attendance Crisis?” *The Electronic Journal of e-Learning* Volume **5** Issue 1, pp 1 - 10, available online at www.ejel.org

Beaudoin, M (2002) Learning or Lurking? Tracking the “invisible” online student, *The Internet and Higher Education* **5** pp147-155

Burd, E and Hodgson, B. (2005) Attendance and Attainment Revisited, *6th Annual Conference of the ICS HE Academy***,** 30th -1st September, University of York (2005).

Colby J., (2004) Attendance and Attainment*, 5th ICS-LTSN. Annual Conference***,** 31st-2nd September, University of Ulster.

Dougiamas, M. (2007) Welcome to Moodle, [online] http://moodle.org

Kirkpatrick, G. (2005) Online ‘chat’ facilities as pedagogic tools, *Active learning in Higher Education*, **6** (5) 145-159

Lally, V. (2002) Elaborating collaborative interactions in networked learning: a multi-method approach, in: V. Lally (Ed.) *Proceedings of the Networked Learning Conference 2002*, University of Sheffield.

Paananen, M. and Simpson, M.(2007), An Exploratory study of the Effects of Web-based Technology on Student Performance, *Second International Blended Learning Conference*, University of Hertfordshire.

Pozzi, F., Manca, S., Persico, D. and Sarti, L. (2007), 'A general framework for tracking and analysing learning processes in computer-supported collaborative learning environments', *Innovations in Education and Teaching International*, **44**:2, 169 - 179

Wang, M., (2004) Correlational Analysis of Student Visibility And Performance in Online Learning. *Journal of Asynchronous Learning Networks* **8** (4)

Woodfield, R., Novell, S. E., and Solomon, L.,(2005) Gender and mode of assessment at university: should we assume female students are better suited to coursework and males to unseen examinations? *Assessment & Evaluation in Higher Education* (**30**), 1, 35 - 50