Using predictive analytics to support students: Case study 1

Alison Gilmour, Avinash Boroowa and Christothea Herodotou (The Open University)*

1. Introduction

The Open University in Scotland (OUiS) has a strategic commitment to improving student retention and progression which is recognised as a more marked issue in distance education than campus-based settings (Simpson, 2013). During academic year 2016-2017 a collaborative pilot project was developed by OUiS staff working with colleagues in other units of the Open University (OU) premised on exploring the potential of Predictive Learning Analytics (PLA), namely the use of predictive modelling and student probabilities, to target retention interventions to improve student retention. This paper details the pilot context, intervention design, and headline outcomes and is purposefully descriptive.

2. Context: retention in distance education and the potential of predictive learning analytics

As an open access distance learning institution, the OU has a strong commitment to widening access accompanied by a strategic focus on retention. The OU's 'Students First' strategy focuses attention on supporting student success and marries with the OUiS ongoing commitment to retention, articulated in the OUiS Outcome Agreement with the Scottish Funding Council (2017). OUiS engagement with the Quality Assurance Agency Scotland managed Enhancement Theme of Student Transitions resulted in a pilot project designed to explore targeting of retention activity.

Within distance learning, the challenge of identifying students experiencing difficulty and offering support is perhaps more so than in a campus-based university, where issues may manifest more evidently in participation, potentially leading to earlier detection. Simpson's work has highlighted the challenges for the OU in module-based retention as well as qualification achievement, and identifies what he labels as a 'distance education deficit' (Simpson, 2013, 105). Key to his work is exploring factors impacting on retention; identifying learner motivation as a significant determining factor. However, his work also refers to retention theorist Tinto, who draws our attention to the interactions students have with institutional systems and processes, and the importance of ensuring there are not opportunities for student 'early exit'. Minimising opportunities for early exit is based on the need for support to ensure student persistence (1993 cited in Simpson, 2012). Indeed, Moore's (1997) notion of 'transactional distance' has some relevance here in signposting the significance in distance education of both the physical and psychological space between students and their tutor/ lecturer, and the significance of communication as one element influencing transactional distance. In the context of retention, narrowing the gap or transactional distance between the student and the institution, potentially through communication, and minimising the opportunities for students to exit prematurely from their studies, are areas of focus within OUiS retention activities.

In the OU, student support is organised within a curriculum aligned Student Support Team (SST) model. SSTs utilise a framework for proactive student support known as Model of Integrated Learning and Learner Support (MILLS). Within MILLS, SSTs identify students at risk of early exit from their studies (hereafter referred to as 'at risk') using factors such as previous levels of education, and undertake proactive contact with these students at designated points within the MILLS framework. Such contact can be automated emails, or direct emails or calls from either the student's tutor or the SST. Students at the OUiS are split across several curriculum aligned SSTs but due to an internal policy known as Dual Affiliation (to both their curriculum area and their nation of residence) students resident in Scotland will also receive additional retention interventions organised by OUiS.

Retention interventions in OUiS are co-ordinated by the Retention Action Group, and between 2011-15 a number of module-focused retention projects focused on: a) exploring whether particular variables could be used to identify students more at risk and targeting support; and b) making use of proactive student support. Exploring the former, there were questions over appropriateness and value of using single variables in a diverse student body, and which did not account for developments in student behaviour over the course of a module. Furthermore, with the latter, there was strong support amongst student support staff and associate lecturers that proactive support was valuable in reaching out to students who may be experiencing difficulty. However, proactive support is resource intensive for the institution and results in difficult questions about to whom the institution could and should offer this type of support. Both of these issues led to the question of whether retention activities could be better targeted, and coupled with a growing awareness of the potential of learning analytics to identify struggling students earlier and to inform retention initiatives, this led to the exploration of the use of predictive modelling and the design of the pilot initiative.

Within our work we used Calvert's definition of predictive analytics as, 'forecasting future outcomes and behaviours, therefore predictive analytics is about forecasting student behaviour. The forecasts are achieved by extracting and processing information from data routinely collected during a business process.' (Calvert, 2014, p. 161) The development of learning analytics encompasses predictive learning analytics and has allowed higher education institutions to reconsider the potential of data on a different scale than previously, through utilising data currently captured within university systems. For example, Arnold and Pistilli argue that key within retention is supporting 'academic integration' of students and they advocate the use of learning analytics incorporating both demographic and behavioural data as rich use of evidence to support student persistence. For them, the application of learning analytics data produces 'actionable intelligence', whereby identifying students who are most at risk of not continuing with their studies should allow for early intervention by the institution (2012, p. 267).

However, whilst there is a growing body of literature citing the potential of learning analytics and specifically the use of predictive learning analytics in the early identification of students, much of the focus within the field has been on the development of predictive models (Herodotou et al., 2017; Calvert, 2014). There has been more limited application and evaluation of predictive learning analytics in the design of retention interventions and consideration of what this means in terms of institutional approaches to retention (Sclater, 2017; Ferguson and Clow, 2017; Dawson et al., 2017; Seidel and Kutieleh, 2017). As part of our strategic focus on retention, OU staff in the Strategy and Information Office and Learning and Teaching Innovation, were exploring the potential use of predictive learning analytics within a project called the Early Alert Indicators Project and coupled with the Enhancement Theme on Student Transitions this provided the impetus to explore the use of predictive analytics as an alternative way of identifying students 'at risk' that contributed to the evolution of existing practices. We were guided by the objective: to improve the retention rate of 'at risk' students as defined by the Student Probability Model (SPM).

3. The Student Probability Model

The Student Probability Model (SPM) produces predictions of whether an individual student will reach specific milestones (different points in a module presentation or between modules) such as completing and passing a module, or returning in the next academic year. The SPM was initially developed by Calvert (2014) and developed over time with large-scale uptake within the OU. Predictions or probabilities in the SPM are based on models generated through logistic regression of a set of 30 explanatory variables. Not all factors are significant in predicting to or from every milestone, but instead a selection are used. As can be seen from Table 1, these variables are grouped by: student factors, module study behaviour, students' previous study, students' previous progress, and module and qualification. Student Probabilities are available from before module start and are updated at selected module milestones, adding new relevant factors as the module progresses - such as virtual learning environment engagement, assignment submissions, number of assignments submitted - and with this comes increasing accuracy of the predictions to determine whether a student will be present at each milestone.

TABLE 1.

Student Probabilities Model	
Student Factors	Demographic factors including Disability, Occupation, Index of Multiple Deprivation.
Module study behaviour	Total numbers of credits being studied, late registration, proportion of assignments submitted to date.
Student's previous study	Credit transfer, previous educational qualification, sponsored or not
Student's previous progress at the OU	OU credit already achieved, number of previous OU passes, withdrawals or fails
Module within a qualification	Current module, study intensity of module, number of assignments due by milestone.

4. Phase 1: The value of a Student Probability Model

The SPM had been used previously with cohorts in other parts of the OU but the project team was interested in the usefulness of student probabilities compared to single variables used previously by OUiS to determine students 'at risk'. We had questions around the suitability of the SPM for use with a Scottish cohort of students, the accuracy of the SPM in predicting non-completion, and how the SPM prediction of non-completions compared with using single variables. In early 2016, Phase 1 of the project tested the application of the SPM with OUiS students, by comparing student probabilities with the actual performance of students studying modules between October 2014 and June 2015.

When comparing single variable selection criteria to using the SPM, Table 2 indicates that of the 'at risk' students identified using student probabilities, a higher percentage of those identified went on to not complete, compared to using other means of selecting potential non-completers. Therefore, from this evidence we felt that using student probabilities is a more reliable method of identifying students 'at risk'.

TABLE 2.
Analysis of students in Scotland studying modules from October 2014-June 2015

Selection Method	**Number in section**	**% of selection who didn't complete**	**Number of non-completers identified**
Student Probabilities at Day 14	**1,968**	**55%**	**1,088**
No formal qualifications	**295**	**42%**	**124**
Progress alert or restricted	**459**	**55%**	**252**
More than 120 credits	**132**	**34%**	**45**
Low SIMD (bottom quintile)	**1,795**	**41%**	**738**

5. Phase 2: Designing an intervention

A retention intervention was designed where students were selected to receive the intervention based on the SPM. The SPM was run seven weeks before module start. Probabilities were generated when students from a mixture of undergraduate courses and levels across the OUiS reached registered status. Students with probabilities to complete their course between 30-40% were selected for the study (N=410). The decision was taken to focus the intervention on this single probability band (30-40%), was based on the availability of resource, namely staff availability to undertake the proactive calls. In order to evaluate the intervention, the decision was taken to take a randomised control trial approach: students were randomly allocated in control (N=200) and intervention groups (N=210). The control group received the standard specialist subject intervention by SST.

Each student in the intervention group received a text message (N=210) to raise their awareness that they were going to receive a call from the University (N=91 students reached by telephone) and after two attempts at calling students, any students that could not be reached on the telephone received an email (N=119). The rationale for the notification text message was that this was deemed appropriate to raise student awareness and prevent a 'sudden call'. The aim of the call was: a) to raise student awareness of sources of support; and b) to ensure that students 'feel equipped' with the information they need to begin their studies. The intervention took place over a three-week period.

In light of the randomised control trial approach and the use of student data, the project team consulted with the OU's Human Research Ethics Committee and Student Research Project Panel. There was agreement that as students in the control group had access to the standard range of support in the University, the benefits of undertaking a pilot with a control group using experimental methodologies outweighed the risks to students. The pilot is aligned with the University's policy on ethical use of student data for learning analytics and is designed to provide better evidence on how to improve the support made available to registered students.

6. Phase 2 intervention outcomes

When considering the outcome of the intervention, as a number of our students studied more than one module continuously, rather than using head count in our analysis we considered the impact across a number of modules. The figures are presented in Table 3:

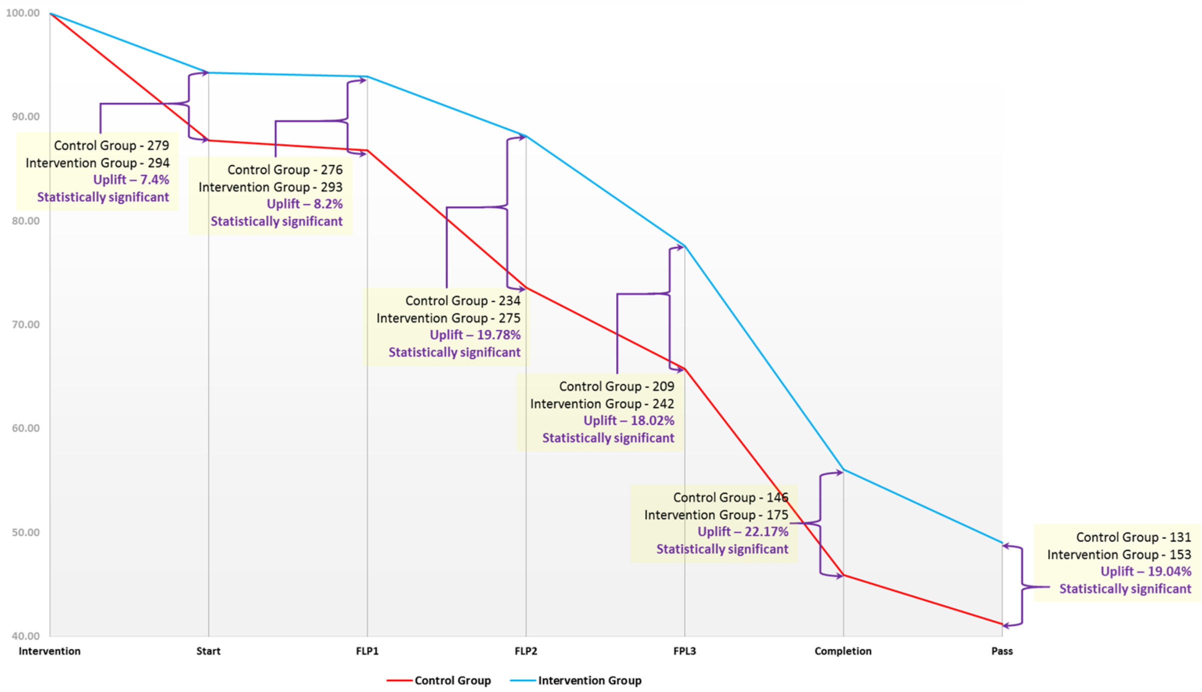
TABLE 3.

Intervention Contact	Number in module count	%
SMS Sent	312	100.00%
Call attempted	305	97.80%
Call successful	132	42.30%
Sent email	172	55.10%

A measure of success within our evaluation was student persistence and retention in the control compared to the intervention group, measured by the presence of students at identified milestones. We ran the SPM at each of the milestones for students studying modules from October 2016 through to June 2017: percentage of student registrations at the start of a module, registrations at 14 days after module start, 3 months after module start, 6 months after course start, and at module completion. Data from as early as the second milestone suggested that the intervention had a positive effect where students in the intervention group indicated statistically significantly lower withdrawal rates that the control group. Statistically significant differences were observed for the intervention group at all milestones suggesting that the proposed intervention was successful and more likely helped students remain engaged and progress through their studies (see Diagram A).

DIAGRAM A.

	Start	14 Days	3 Months	6 Months	Completion	Pass
Gross (Control – 318)	279	276	234	209	146	131
Gross (Intervention – 312)	294	293	275	242	175	153
Net (uplift)*	21 (7.4%)	23 (8.2%)	46 (19.78%)	38 (18.02%)	32 (22.17%)	25 (19.04%)



7. Conclusion

This paper details the context, design and initial outcomes of pilot activity that made use of PLA to identify students with a low probability of completing their modules and offered proactive support, with a randomised control trial approach used to build an evidence base. This paper highlights the potential use of PLA to design retention activities focused on early identification of students at risk, in ways which compliments existing practice. PLA intervention at OUiS could be viewed as an alternative way of choosing and supporting students that aligned well and contributed to the evolution of existing practices used to identify students at risk. Indeed, Sclater has highlighted that ‘Learning analytics requires bringing people with high levels of technical expertise together with others who understand pedagogy and educational processes.’ (2017, p. 16) Within the pilot project, a collaborative team was able to build on previous and existing student support and retention activities, but make use of PLA in order to make decisions about identifying students ‘at risk’ in order to target resource intensive early intervention. It is likely that the close similarity between the proposed PLA intervention and existing practices of identifying and intervening with students more likely favoured successful adoption and implementation by support staff.

For the project team, exploring how PLA could improve current practice and strengthen our evidence around the potential use of PLA was key. The project has provided insight as to the usefulness of the SPM in identifying students at risk, as well as on the design of the intervention. We continue our investigation into initial outcomes and a development of the pilot is underway with a 2017-18 cohort of students. This work supports previous OUiS retention work that proactive support seems to have positive retention and progression effects with students, although a repeat of this pilot, and extension of the evaluation exploring different types of data for analysis, would strengthen the evidence base and our understanding of the impact of PLA.

**The paper authors want to acknowledge the involvement of a broad team of staff working collaboratively to shape the nature of the intervention and therefore the content of the paper. Significantly: Hannah Jones, Lucy Macleod, Galina Naydenova and Rebecca Ward.*

<u>References</u>

Arnold, K. and Pistilli, M. (2012) 'Course Signals at Purdue: Using Learning Analytics to Increase Student Success', *LAK '12 Proceedings of the Second International Learning Analytics and Knowledge Conference*, ACM International Conference Proceeding Series, ACM, New York, pp. 267-270 [Online]. DOI: 10.1145/2330601.2330666 (Accessed 10 May 2018).

Simpson, O. (2013) 'Student retention in distance education: are we failing our students?' *Open Learning: The Journal of Open, Distance and e-Learning*, vol. 28, no, 2, pp. 105–119 [Online]. DOI: 10.1080/02680513.2013.847363 (Accessed 10 May 2018).

Simpson, O. (2012) *Supporting Students for Success in Online and Distance Education* (Third Edition), New York, Routledge.

Calvert, C. (2014) 'Developing a model and applications for probabilities of student success: a case study of predictive analytics', *Open Learning: The Journal of Open, Distance and e-Learning*, vol. 29, no. 2, pp. 160-173 [Online]. DOI: 10.1080/02680513.2014.931805 (Accessed 10 May 2018).

Ferguson, R. and Clow, D. (2017) 'Where is the evidence? A call to action for learning analytics', *LAK '17 Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, ACM International Conference Proceeding Series, ACM, New York, pp. 56–65 [Online]. DOI: 10.1145/3027385.3027396

Dawson, S., Jovanovic, J., Gasevic, D. & Pardo, A. (2017) 'From prediction to impact: Evaluation of a learning analytics retention program', *LAK '17 Proceedings of the 7th International Conference on Learning Analytics and Knowledge*, ACM International Conference Proceeding Series, ACM, New York, pp. 474-478. DOI: 10.1145/3027385.3027405

Seidel, E. and Kutieleh, S. (2017) 'Using predictive analytics to target and improve first year student attrition', *Australian Journal of Education*, vol. 61, no. 2, pp. 200-218 [Online]. DOI: 10.1177/0004944117712310 (Accessed 10 May 2018).

Herodotou, C., Heider, S. and Rientes, B. (2017) 'Implementing Randomised Control trials in Open and Distance Learning', *Open Learning: The Journal of Open, Distance and e-Learning*, vol. 32, no. 2, pp. 147-162 [Online]. DOI: 10.1080/02680513.2017.1316188 (Accessed 10 May 2018).

Moore, M. (1997) 'Theory of transactional distance' in Keegan, D. (ed) *Theoretical Principles of Distance Education*, London, Routledge, pp. 22-38 [Online]. Available at http://www.c3l.uni-oldenburg.de/cde/found/moore93.pdf (Accessed 20 May 2018).

Sclater, N. (2017) *Learning Analytics Explained*, New York, Routledge.

The Open University in Scotland (2017) *The Open University in Scotland Outcome Agreement* [Online], Edinburgh, The Open University in Scotland. Available at http://www.sfc.ac.uk/web/FILES/outcome-agreements-1718/open-university-outcome-agreement-2017-18.pdf (Accessed 12 May 2018).