

Widening the net to reduce the debt: Reducing student debt by increasing identification of completely disengaged students

Neil Van Der Ploeg	Kelly Linden	Ben Hicks	Prue Gonzalez
Charles Sturt University Australia	Charles Sturt University Australia	Charles Sturt University Australia	Charles Sturt University Australia

Student Retention and Attrition guidelines are part of the Federal Government’s performance based funding framework. One of the recommendations from the Higher Education Standards Panel review is to consider changing students’ enrolment prior to census date when a certain level of engagement is not met. This study investigates this recommendation by trialing and testing a model to see if completely disengaged students are able to be retrospectively identified as at risk of failing all subjects. Using learning analytics alone to create a predictive model at scale proved to be very difficult. When applied to session 1 of 2019, even the strictest criteria included five false positives out of 17 identified students. There is promise, however, that a hybrid model of learning analytics with additional oversight from teaching staff could be a solution, but this needs further research.

Keywords: Learning analytics, retention, engagement, attrition, student support

Introduction

Student progression, attrition and completion rates matter not only to Universities but also to the Australian Government and the broader community. In 2020, it matters more than ever to Australian universities given that the Federal Government’s new performance-based funding policy comes into effect. Attrition is a metric that is part of the new legislation:

The Department of Education and Training counts as attrition a commencing student who reached their first census date (which is at least 20 percent of the way through the semester) but is not enrolled in the next year without completing (Cherastidtham & Norton, 2018, p. 5).

Although attrition is the actual metric in the Government’s policy, it has practical limitations because the reporting lags for at least 12 months after the student enrolls. Subject progress rates are often used as an imperfect, but strongly related proxy for attrition when working within a narrower window of time. This is the rough rule of thumb: improving progress rates today, will improve attrition rates tomorrow.

Challenges to retention are multi-factorial, hence the wide range of recommendations made in the Department of Education and Training’s report (2017) on improving retention. There have been varying degrees of research that can guide the implementation of some of these recommendations. Our team has previously piloted pre-census targeted student support based on low engagement. The low engagement was defined through a missed early assessment item and the support was provided through phone contact by the outreach team who gave immediate advice and if appropriate referred students to existing support services. Their role is analogous to a general practitioner in healthcare. This pilot appeared to improve progress rates across 25 first year subjects (Linden & Webster, 2019). While the student support approach is effective in re-engaging students with their studies and increasing progress rates, it does not work for all students (Linden & Webster, 2019). What emerged from the pilot and recent unpublished data was that many students are unresponsive to contact by the University. Some of these students have been described as “ghost” students (Stephenson, 2019, p. 1): they never engage in studies and fail all of their subjects without submitting any assessment items. Clearly, there needs to be a different strategy for the students who never engage, or engage at a level that does not constitute a genuine attempt at study. Indeed, the Department of Education and Training (2017) has made a recommendation that targets these students: “institutions should automatically review the enrolment of all students who have not engaged in their studies to an agreed level by the census date” (p. 9). However, exactly what this “agreed level”

should be is not self-evident. Furthermore, there has been little discussion or research within the higher education sector to provide guidance. Given these measures must be applicable at scale, learning analytics are an obvious tool that universities will reach for.

Learning analytics to improve retention

Twenty years ago, when advances in digital technology were integrated into tertiary education, there were predictions that it would lead to “actionable intelligence” (Campbell et al., 2007, p. 42). As early as 1999, universities have designed models to identify students who are at risk of failing with the intention of providing targeted support (in a variety of possible forms) to these students (Campbell et al., 2007). Through this process of targeted support, the hope is that universities can increase student retention and reduce attrition. Most of the learning analytics research is underpinned by this philosophical framework.

The creation of accurate at risk models has proved to be very challenging. Over the past decade, universities have tried several different approaches to designing at risk models. Broadly speaking, models attempt to balance the demographic, academic and activity data of students to some degree, however, very different approaches exist concerning model complexity (Wolff et al., 2014; Kuzilek et al., 2015; Jayaprakash et al., 2014; Foster & Siddle, 2019; Akçapınar et al., 2019). Despite years of research and development of a complex model at The Open University, the predictive capacity was far from perfect, especially when attempting to make predictions early in the session (Kuzilek et al., 2015). Another approach is to acknowledge the inherent complexities of the system in predicting student retention (Forsman et al., 2014) and use a simplified model. Once again however, the predictive capacity of simplified models is modest at best at early points in the session, and improves over the duration of a session (Foster & Siddle, 2019). This presents an even greater challenge given that the decision to defer a student must be made prior to census date, with limited time and therefore data to feed into the model.

In order to save as many students as possible from accumulating unnecessary debt without adversely affecting attainment, this paper sets out to answer the following question by looking at retrospective data: how many students headed for complete academic failure can be identified using learning analytics data available at census date, without also identifying any students who were making a genuine attempt at study? This will necessarily require an at risk model with much tighter conditions than what has been previously published in the literature. If such a model could be designed, universities could consider using it to defer or cancel the enrolments of completely disengaged students without interfering with students who are on track for some success.

Methods

Educational context

Charles Sturt University is a large Australian tertiary institution that spans multiple campuses, predominantly in regional areas of New South Wales. It is one of the largest providers of online and blended learning in Australia and has approximately 40,000 students. Undergraduate students typically attend courses of 3 to 4 years duration, delivered over the first two trimesters of each year. These will be referred to as session 1, and session 2. The third trimester (session 3) is typically used for additional flexibility and accelerated learning and is not part of this study. This paper focuses on session 1 of 2019. The census date at Charles Sturt University is on the Friday of the 4th week of session 1. In Australia, the census date at a University is when domestic students must decide if they wish to commit to studies for the given session. Once past the census date, students are liable to pay course fees and will have subject grades recorded on their permanent academic record.

Exploratory analysis

This paper uses an exploratory, retrospective analysis that aims to investigate models that could predict if students are at risk of complete failure. To do this, different models are tested by retrospectively applying them to student data from session 1 of 2019 and observing the eventual academic outcomes of those students.

Subject and student selection

Commencing, undergraduate students from all three Faculties of the University were included in the study (Faculty of Science, Faculty of Arts and Education and the Faculty of Business, Justice and Behavioural Studies). Commencing students were selected because if they do not remain enrolled beyond the census date in

their first session, they do not count as attrition.

Student contact

In 2019, the Retention and Engagement Team identified disengaged students enrolled in almost all undergraduate courses in the critical pre-census window. In week 2 an email was sent to over 1000 students who had not logged into the Learning Management System (LMS) since the commencement of session. The email was a friendly reminder that the session had commenced. In weeks 3 and 4, 1193 students were contacted by phone due to either non submission of an early assessment item, or low activity in the LMS. After 2 failed call attempts, an email was sent to students with information about the financial implications of remaining in a subject without engaging, and support to unenroll if a student had changed their mind. Thus it was anticipated that all disengaged, commencing, undergraduate students received at least one communication regarding retention.

Exploratory Findings

Creating a model that could predict if students will fail all subjects

Four activity metrics over the first 4 weeks of session were used to generate the model: (1) number of accesses/logins, (2) number of views, (3) items viewed and (4) the time spent in the LMS. This model was applied retrospectively to almost 5000 students.

Attempting to use the four activity metrics to predict completely disengaged students yielded unsatisfactory results, particularly when high specificity of the model is paramount. Splitting the data into 60% training and 40% testing and applying a logistic regression model, the resulting ROC curve (Figure 1) is only slightly above the diagonal, showing that the logistic regression is only slightly better than random guessing (Sing et al., 2005). Notably, the curve has particularly poor performance when trying to keep the false positive rate low, which is critical when flagging a student as completely disengaged. The activity variables were trialled in isolation and also combined. Adding more activity variables to the model did not appear to improve performance. A different approach was needed.

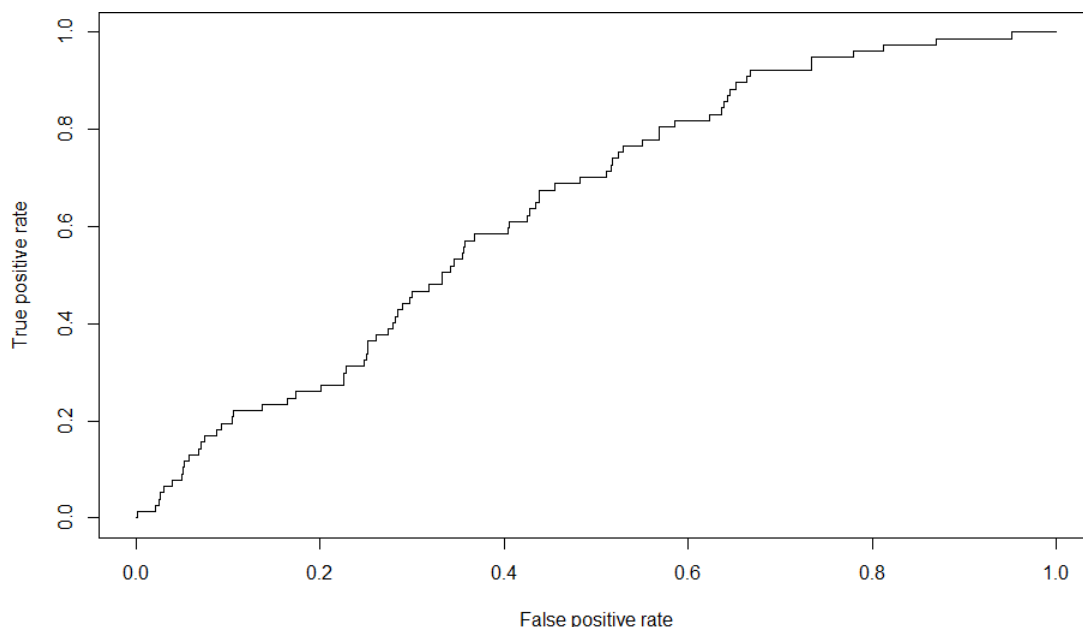


Figure 1. ROC Curve of logistic regression model, predicting failing all subjects based on LMS activity metrics; logins, views, items viewed and minutes

Back to basics: No LMS activity

We next explored whether the date of last LMS access would provide a more accurate method of identifying

completely disengaged students, remembering that all disengaged students should have been sent an email in week 2 and received phone or email contact in week 4. In addition, they were likely to have missed at least one pre-census assessment item.

Rather than summarising logins as an aggregate per subject per week over the first four weeks, students were identified by time since last LMS login. Given the low numbers of students, they were merged into three groups, depending on which week they last accessed the LMS. This created three levels of criteria, progressively becoming more inclusive (or less strict). The hope was to find the criteria that could identify as many ‘failing all’ students as possible, without including any students who went on to pass a subject. The most strict criteria only identified students with absolutely no LMS access from weeks 1 to 4, meaning that these students had a period of four weeks leading up to census date without a single log in. The next criteria identified students who had no LMS activity during weeks 2 to 4, meaning that these students had a period of three weeks leading up to census date without any activity. The final criteria identified students who had a period of inactivity of only two weeks prior to census.

The results were surprising. The simple model based on LMS access alone was unable to accurately identify students who went on to fail all subjects without also including students who went on to pass at least one subject. As can be seen in Table 1, even with the tightest criteria there were students who passed at least one subject, who at census date, had not logged into the LMS on a single occasion for any of their subjects. As the criteria were widened this error was exacerbated. Absence from the LMS for 2 weeks included a large number of students (n=22) who would still go on to pass at least one subject. Using learning analytics at scale to exclusively identify disengaged students at risk of complete failure seems to be comparably difficult to predicting students at risk of some failure (Wolff et al., 2014; Kuzilek et al., 2015; Jayaprakash et al., 2014; Foster & Siddle, 2019; Akçapınar et al., 2019).

Table 1. Number and outcome of students identified as ‘at risk of failing everything’ by a simple model using learning analytics alone, without any subjects/courses removed at census date.

Learning analytics alone			
<i>Time since last LMS login</i>	<i>Number of identified students</i>	<i>Passed something (False positive)</i>	<i>Failed everything (True positive)</i>
4 weeks	17	5	12
3 weeks	26	9	17
2 weeks	56	22	34

When looking at the subjects and courses of these students however, it was clear that certain subjects were over-represented in the students who were identified as completely disengaged, but went on to pass (false positives). Three students in the strictest criteria (no LMS access in the 4 weeks leading up to census date) were all enrolled in the same subject and course. Consulting with academic staff who had knowledge of the course providing false positives indicated that these subjects have an atypical delivery of subject content. In other words, the academics who understood the delivery of these subjects did not expect LMS access to be a valid indicator of complete disengagement. This is an example of the challenges of applying at risk models at scale using learning analytics alone; occasional subject and course variation can undermine critical assumptions of the model. In this case the assumption was as follows: if students do not access the LMS at all and do not respond to several contact attempts by the student outreach team, then they must be completely disengaged and on track for failure. Clearly, this assumption cannot be made in all subjects. After removing the identified atypical online subjects/courses through dialogue with academic staff, the results were very different (Table 2.).

Table 2. Number and outcome of students identified as ‘at-risk of failing everything’ by a simple model using learning analytics with 5 subjects removed on the advice of academic staff.

Learning analytics with excluded subjects/courses			
<i>Time since last LMS login</i>	<i>Number of identified students</i>	<i>Passed something (False positive)</i>	<i>Failed everything (True positive)</i>
4 weeks	8	0	8
3 weeks	12	0	12
2 weeks	24	3	21

The number of true positives dropped slightly, but more importantly, the number of false positives also dropped. Overall, the specificity was greatly improved. The net could be widened to include students who had not accessed the LMS for 3 weeks leading up to the census date without falsely identifying any students who went on to pass. Had this model been applied at census day of session 1 in 2019, 12 completely disengaged students could have been identified and deferred, avoiding significant student debt.

The Department of Education and Training's recommendation (2017) to vary students' enrolment prior to census if an agreed level of engagement has not been met is more complicated than one may initially think. Without academic input, even when applying the strictest definition of disengagement using LMS access data (no LMS access prior to census date), some students were captured who did achieve some degree of success.

This paper has applied a retrospective analysis and therefore has inherent limitations. Firstly, had the students known they would be deferred if they were absent from the LMS, the more engaged students may have altered their behaviour and accessed it. Regardless, if more data could be added into the model it is possible that it could be further strengthened. Failing to submit an early assessment item could be such a data point. Wolff et al. (2014) found that failure to pass the first assessment item was a key predictor of academic success. Failing through non submission or receiving insufficient marks led to a "high probability" of overall failure. This supports soon to be published data from the Charles Sturt University Retention team that indicates students who do not submit a pre-census early assessment item are far less likely to pass the subject. More research is needed in this area to inform university policy changes.

Conclusion

This study attempted to create a simple at risk model using learning analytics that could, at census date, exclusively identify students who were disengaged and on track for complete failure. The use of simple analytics alone, at scale, was not able to do this. This was the case even when commencing, online students were completely absent from the LMS for four weeks leading up to the census date. A potential approach emerging in this paper is to communicate with the academic staff, and change the model based on expert course knowledge. This would be a hybrid model that relies on data from learning analytics and data directly from teaching staff. As we head into a post-pandemic era, the damaging financial effects will impact both students and governments alike. Saving students unnecessary debt while not inhibiting their chances of success is increasingly important but no less difficult.

References

- Akçapınar, G., Hasnine, M. N., Majumdar, R., Flanagan, B., & Ogata, H. (2019). *Developing an early-warning system for spotting at-risk students by using eBook interaction logs*. *Smart Learn. Environ.* 6(4). <https://doi.org/10.1186/s40561-019-0083-4>
- Campbell, J., DeBlois, P., & Oblinger, D. (2007). Academic Analytics: A new tool for a new era. *EDUCAUSEreview*, 42(4),40–57. <https://er.educause.edu/articles/2007/7/academic-analytics-a-new-tool-for-a-new-era>
- Cherastidham, I., & Norton, A. (2018). *University attrition: what helps and what hinders university completion?* Grattan Institute Background Paper No. 2018, November. Retrieved from <https://grattan.edu.au/wp-content/uploads/2018/04/University-attrition-background.pdf>
- Department of Education & Training. (2017). *Improving retention, completion and success in higher education*. Higher Education Standards Panel Discussion Paper, June. Retrieved from https://docs.education.gov.au/system/files/doc/other/final_discussion_paper.pdf
- Foster, F., & Siddle, R. (2019). The effectiveness of learning analytics for identifying at-risk students in higher education. *Assessment & Evaluation in Higher Education*, 6 November. <https://doi.org/10.1080/02602938.2019.1682118>
- Forsman, J., Linder, C., Moll, R., Fraser, D., & Andersson, S. (2014). A new approach to modelling student retention through an application of complexity thinking. *Studies in Higher Education*, 39(1), 68–86.
- Jayaprakash, S. M., Moody, E., Lauría, E., Regan, J., & Baron, J. D. (2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, 1(1), 6–47.
- Kuzilek, J., Hlosta, M., Herrmannova, D., Zdrahal, Z., Vaclavek, J., & Wolff, A. (2015). OU Analyse: Analysing at-Risk Students at The Open University. *Learning Analytics Review LAK*, 15(1), 1–16.
- Linden, K., & Webster, L. (2019). *Back to Basics: combining analytics and early assessment with personalised contact to improve student progress* [paper presentation]. 36th International Conference on Innovation,

- Practice and Research in the Use of Educational Technologies in Tertiary Education (ASCILITE). Singapore.
- Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. (2005). ROC: visualizing classifier performance in R. *Bioinformatics*, 21(20), 3940–3941.
- Stephenson, B. (2019). Universities must exorcise their ghost students. *Times Higher Education*, <https://www.timeshighereducation.com/opinion/universities-must-exorcise-their-ghost-students>
- Wolff, A., Zdrahal, Z., Herrmannova, D., Kuzilek, J., & Hlosta, M. (2014). *Developing Predictive Models for Early Detection of At-Risk Students on Distance Learning Modules* [Paper presentation]. Machine Learning and Learning Analytics Workshop at The 4th International Conference on Learning Analytics and Knowledge (LAK14), 24–28 March. Indianapolis, Indiana, USA.

van der Ploeg, N., Linden, K., Hicks, B. & Gonzalez, P. (2020). Widening the net to reduce the debt: Reducing student debt by increasing identification of completely disengaged students. In S. Gregory, S. Warburton, & M. Parkes (Eds.), *ASCILITE's First Virtual Conference*. Proceedings ASCILITE 2020 in Armidale (pp. 54–59). <https://doi.org/10.14742/ascilite2020.0125>

Note: All published papers are refereed, having undergone a double-blind peer-review process. The author(s) assign a Creative Commons by attribution licence enabling others to distribute, remix, tweak, and build upon their work, even commercially, as long as credit is given to the author(s) for the original creation.

© van der Ploeg, N., Linden, K., Hicks, B. & Gonzalez, P. 2020