

Linking early alert systems and student retention: a survival analysis approach

Renato Villano¹ · Scott Harrison¹ · Grace Lynch¹ · George Chen¹

Published online: 3 March 2018

© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract Higher education institutions are increasingly seeking technological solutions to not only enhance the learning environment but also support students. In this study, we explored the case of an early alert system (EAS) at a regional university engaged in both on-campus and online teaching. Using a total of 16,142 observations captured between 2011 and 2013, we examined the relationship between EAS and the student retention rate. The results indicate that when controlling for demographic, institution, student performance and workload variables, the EAS is able to identify students who have a significantly higher risk of discontinuing from their studies. This implies that early intervention strategies are effective in addressing student retention, and thus an EAS is able to provide actionable information to the student support team.

Keywords Early alert systems · Student retention · Learning analytics · Survival analysis

Introduction

Student retention has been a significant area of research within universities for a long time. It is a complex issue, where the learning environment is a myriad of interactions between students, academics and administrators within the higher education system. Increasingly, various

✉ Renato Villano
rvillan2@une.edu.au

Scott Harrison
sharri52@une.edu.au

Grace Lynch
glynch2@une.edu.au

George Chen
gchen2@une.edu.au

¹ UNE Business School, University of New England, Armidale, NSW, Australia

institutions have developed and implemented early alert systems (EAS) as part of an early alert programme (EAP). The distinction between EAP and EAS is that the EAP is a type of initiative that universities can undertake to promote early intervention and support of students, whereas EAS refers to the actual IT programmes and algorithms used in identifying which students may need assistance. EAS are designed to aid student support teams in identifying students at risk in order to provide proactive student supports.

While some studies have examined the effectiveness of EAS, one limitation has been the treatment of temporal effects. Where temporal modelling has been applied (DesJardins 2003; Ishitani 2006), the analysis was limited to analysing effects between initial and subsequent teaching periods. This resulted in limited understanding of how the risk of discontinuing can vary within teaching periods. This study provides a detailed time series analysis of retention that is envisioned to increase the understanding of risk of discontinuing, both during and between teaching periods with greater accuracy than previously possible. Using 156 weeks of data, our aims are to establish (i) the relationship between demographic, institution, student performance and workload factors and student retention, (ii) the effects of these factors over time and (iii) the relationship between EAS and student retention.

The paper is organised as follows: ‘[Research context](#)’ provides an overview of the relevant research context focusing on the literature on EAS, student retention using survival analysis and the evaluation of learning analytics initiatives; ‘[Method of analysis](#)’ presents the methods of analysis, including a detailed discussion of the EAS algorithm; and ‘[Results](#)’ presents the results of the analysis followed by summary and concluding comments in ‘[Summary and conclusion](#)’.

Research context

The area of student retention research is a mature and well-established body of knowledge, focusing on the problems of retention and attrition from both theoretical and practical viewpoints. For the purpose of this study, student retention is defined as students who remain enrolled at university; they do not discontinue through formal administrative processes nor do they lapse their enrolment where the student fails to undertake any units of study which count towards a degree.

Early alert systems

Growing awareness of learning analytics corresponds to increasing interests in using EAS to enhance student retention. EAS are designed for the early detection of at-risk behaviour, with the objective to support initiatives aimed at improving student retention, reducing student attrition and/or supporting students at risk of disengaging. A common feature of the initiatives is proactive contact with students upon the EAS identifying someone as being at risk of disengaging, discontinuing or failing a unit (Nelson and Creagh 2013). Several universities have developed systems for identifying students at risk, such as the Open Academic Analytics Initiative (OOAI) at the Marist College (Jayaprakash et al. 2014, p. 7) or the Course Signals System at Purdue University. Course Signals ‘detects early warning signs and provides interventions to students who may not be performing to the best of their abilities before they reach a critical point’ (Purdue University Information Technology 2013). The programme implements many aspects of the underlying student retention theory, where ‘signals combines

demographic information with online engagement’, allowing the programme to factor in some measures of academic integration (Straumsheim 2013). Other commercial systems are now available in this domain (Drop out Detective, Retention Centre in Blackboard). However, in most cases, interest in developing an EAS failed to manifest into a system deployed to improve student outcomes (Jayaprakash et al. 2014, p. 11). Implementing an early alert initiative ‘requires strong leadership and awareness to instil a coherent vision and strategy and to navigate the complexities and resistance to change’ (Siemens et al. 2013, p. 29).

To make an early detection of at-risk behaviour, EAS relies on different sources of data for supporting analysis. One source of data is the learning management system (LMS), which has been found to be a reliable source of information for students’ grades and login data when compared to more traditional sources of student data used for support (Krumm et al. 2014, p. 107). Other EAS data sources include learning tools such as virtual machines (Romero-Zaldivar et al. 2012), data obtained by linking online portals to observations made by trained practitioners (Baker et al. 2010; Baker et al. 2015) and student demographic and aptitude data (Jayaprakash et al. 2014, p. 14). The integration of multiple data sources within the EAS paints a fuller picture of the learning environment to be developed.

For learning analytics as a whole, a central issue is the ethics associated with using student data. Specific ethical concerns of EAS include issues of ‘data ownership, the ethics of surveillance and potential, and the harm to students through labelling’ (Lawson et al. 2016, p. 962). Students have a strong case for owning personally identifiable data when it affects their lives in a personal, professional or academic context (Jones et al. 2014, p. 5). Students should be empowered to decide how their data is utilised by institutions, as institutions may utilise data in ways that are not transparent to the student in allowing informed consent (Lawson et al. 2016, p. 963). While not addressing all ethical concerns, one common feature is the ability of students to either opt-in or opt-out of the systems (Prinsloo and Slade 2016, p. 167), thus empowering students to have some control over how their data is used.

Early alert initiatives can be summarised as a proactive evidence-based approach to connecting students to student support in a timely manner. To be evidence-based, institutions need to make reliable student data available for the implementation of the programmes. To be proactive, the early alert initiatives need to provide actionable data to advisors and support staff. To connect students to support, students need to be contacted directly with options to assist them in making informed decisions about their study and support options. To understand the benefits of EAS, an evaluation is required; we will return to this point in ‘[Programme evaluation](#)’.

The EAS at University of New England (UNE)

The University of New England (UNE) is an established bricks and mortar university that services a large off-campus online student cohort. As such, this makes for an ideal case university, as comparisons can be made between online and off-campus online students, allowing inferences to be drawn on how EAS can work in both settings. A custom-designed EAS called the Automated Wellness Engine (AWE) uses student-level information from a data warehouse to analyse, flag and report students deemed at risk of disengaging from their studies. The data warehouse collects and stores data from eight IT platforms within the institution. This includes unit monitoring reports on teaching, course-level information and university workforce data. The EAS uses 34 triggers to identify at-risk students, with each trigger assigned a positive or negative weight, summated to give a final score per student. A

description of the 34 triggers is contained in Table 1. Each day, 200 students with the highest negative scores are sent an initial email outlining support options, and they could contact the student support team to opt-in to a tailored support programme at their discretion (Leece and Cooper 2011). Specific triggers to the EAS include the alternative entry pathway variable which occurs when the student is admitted on the recommendation of the high school principal, rather than the admission scores. Another key trigger is the in-house tool called ‘e-Motion’ which consists a set of emoticons in the student portal that appears next to each unit the student is currently enrolled in.

The student can select several states to represent how they feel about their studies, including ‘happy’, ‘neutral’, ‘I do not want to say (opt out)’, ‘unhappy’ and ‘very unhappy’. Triggers 11–14 represent e-reserve activities which shows the students’ use of the online library portal as a means of accessing information relevant to studies. Finally, the teacher enabling course is a special course for developing careers in education and for students who do not have skills that align with required skills sets for employment.

Table 1 EAS triggers used to identify at-risk students

Trigger	Description
1	Student admitted through alternate entry pathway
2	Student is Aboriginal or Torres Strait Islander
3	Unit is a currently high attrition unit
4	Unit is a historically high attrition unit
5	Student is a college resident
6	Student registered ‘Happy’ in e-Motion
7	Student registered ‘I do not want to say’ in e-Motion
8	Student registered ‘Neutral’ in e-Motion
9	Student registered ‘Unhappy’ in e-Motion
10	Student registered ‘Very Unhappy’ in e-Motion
11	Student has high e-Reserve usage inactivity (31–40 days)
12	Student has low e-Reserve usage inactivity (10–20 days)
13	Student has medium e-Reserve usage inactivity (21–30 days)
14	Student has very high e-Reserve usage inactivity (41+ days)
15	Student has been granted one to two assignment extension in current teaching period
16	Student has been granted more than two assignment extension in current teaching period
17	Student has submitted one to two assignments late in current teaching period
18	Student has submitted more than two assignments late in current teaching period
19	Student enrolment has involved > double their number of currently enrolled units in current teaching period, post-start of teaching
20	Student has appeared in High Risk Category in a previous teaching period
21	Student is an international student
22	Student has no prior enrolment at UNE
23	Student is enrolled in five or more units in a single teaching period
24	Student was previously enrolled in a pathways enabling course
25	Student has been flagged for contact by the retention team in current teaching period
26	Student has been flagged for contact by the retention team in a previous teaching period
27	Student is carrying over Special Extension of Time (SET) exams from a previous teaching period and is enrolled in current teaching period
28	Student has high portal usage inactivity (31–40 days)
29	Student has low portal usage inactivity (10–20 days)
30	Student has medium portal usage inactivity (21–30 days)
31	Student has very high portal usage inactivity (41+ days)
32	Student was enrolled in the Teacher Enabling course
33	Student received a fail in a unit in a prior teaching period
34	Student received a fail incomplete in a unit in a prior teaching period

From Leece and Cooper (2011)

Some of the variables used by EAS at UNE reflect salient characteristics of the institution. In general, these EAS variables can be summarised into three main categories. The first category is demographic variables, which capture the students' background factors that are independent of their enrolment within the institution. The second category is institutional variables, which capture institutional level data relevant to the students' learning, including information on the course undertaken, tuition fees charged and being on-campus or off-campus online. The third category relates to the student performance and workload variables, which correspond to factors such as assessment information taken throughout a unit of study.

With significant amounts of data available, the EAS at UNE has taken an integral role in helping direct student support efforts. However, the central motivation to this study is understanding the relationship between student retention and the EAS. Therefore, the definition of an at-risk student for this study is a student who ranks in the top 200 students on any given day, based on the 34 criteria used by the EAS. This study focuses on evaluating the EAS efficacy to identify students at risk by linking it to the retention rate.

Programme evaluation

There has been significant focus on the design and implementation of EAS through various initiatives; however, the literature on the evaluation of EAS remains limited. In one study, the evaluation process formed an important dimension to the model developed by Tynan and Buckingham Shum (2013). The model was implemented as part of the Open University (OU UK) strategic analytics investment programme (Ferguson et al. 2015, p. 133).

Evaluation becomes a reflective process used to analyse whether the project achieved its initial objectives. It is important to establish what can be learnt from the process of designing and implementing an EAS and to measure the size of the effect it has on student retention. Several early alert initiatives have reported improvements in retention. For example, Arnold and Pistilli (2012) estimated that the benefits of Purdue's Course Signals improved graduation rates by 21%. However, Caufield (2013b) questioned the results, for not making it clear if the number of courses were controlled for. It was not possible to disaggregate the effects of students taking more Course Signals courses because they persist, compared to students persisting because they are taking more Course Signals courses (Caufield 2013a). Without separating these effects, it is not possible to infer the efficacy of Course Signals in addressing retention.

In a pilot study conducted at the University of Sydney, 'moderate success, particularly among the Business participants, whose failure rate was 50% below that of the BUSS1001 cohort' (Khamis and Kiernan 2013, p. 4) was reported. In the case of Royal Melbourne Institute of Technology's (RMIT) Student Success Program (SSP), it reported that the programme had increased student retention from 57.8 to 64.8% (an extra 40 students). In turn, it was estimated by RMIT that the programme brought in an additional \$1.13 million of revenue, net of the administration costs of the programme (Nelson and Creagh 2013, p. 80). Queensland University of Technology (QUT) reported similar success, finding 'estimated retained income through the retention of an additional 227 students is \$3.75 million for every remaining year of their enrolment' (Nelson and Creagh 2013, p. 72). These estimates show that early alert initiatives have the potential to be prudent investments for institutions implementing EAS.

One major issue in evaluating EAS is over-simplified measures of percentage changes in retention rates that do not accurately reflect the true benefit of the systems. In the case of QUT, it was reported that 'in three out of five units (units 1, 2 and 4), the Student Success Program

(SSP) intervention with at-risk students had a statistically significant impact upon their achievement' (Marrington et al. 2010, p. 2). This indicates that for units 3 and 5, there was no significant improvement in student outcomes. This raises questions over the validity of conclusions drawn on the SSP's effectiveness in the general context, especially if institutions are seeking to implement initiatives on a large scale.

Finally, evaluating EAS is important for justifying the expenditure of university resources on such a programme. Harrison et al. (2016) indicated positive financial benefits to an institution with an EAS. The study used econometric modelling techniques to provide a solid case for the benefits of an EAS. However, their findings can only be validated if the study is replicated in other settings. This study contributes to the field by providing a detailed evaluation of an EAS using time series analysis with a level of temporal resolution previously not available.

Case study: EAS design and conceptual framework

The EAS and course progression framework from UNE is presented in Fig. 1. The flowchart outlines the daily processes students go through from enrolment to outcomes, factoring in student support services and the EAS. Several key events can occur throughout students' enrolment. These events are denoted with X , where a student will either be identified by the EAS (X_1) or not identified by the EAS (X_2). If the student is identified by the EAS, the next event is being contacted by the student support team. If the student was contacted (X_{11}), then the student may opt for targeted student support (X_{111}). The default path is coloured in blue, where a student will not be identified by the EAS and will continue (Y_1) their studies. This creates the circular path that surrounds the EAS system and represents the most common path of students. The EAS identification process occurs daily, implying a circuit following the blue path that takes 1 day.

The framework shows that, on any given day, a student can choose to discontinue their studies. Completion (Y_2) is limited only to those who have satisfied the conditions to be admitted to the award of their course. The remaining outcomes have different meanings for situations students choose when stopping or dropping out of UNE. The first is inactive (Y_3) which captures students who are enrolled in the institution but choose to not enrol in any units of study in a teaching period. This is one of two stopout scenarios, capturing the situation where students have not formally notified the university of their intention to defer studies. The second outcome is intermittent enrolment (Y_4). This is the situation where a planned stopout of studies is taken, formally applying for studies to be deferred. The two remaining outcomes (Y_5 and Y_6) relate to students stopping their studies altogether.

Lapsed (Y_5) refers to the situation where a student has been inactive for at least 2 years. Lapsed captures a student who has stopped their studies but has not administratively discontinued their studies. The final outcome, discontinued (Y_6), is where the proper administrative processes have been followed and the student ceases to study their enrolled course.

A central part of the process outlined in Fig. 1 is the identification of students in need of assistance by the EAS (event X). The identification process utilises 34 triggers reflecting data points collected on students throughout the learning process. Some triggers are static throughout the enrolment of the student (e.g. Aboriginal and/or Torres Strait Islander status, an important variable capturing an academically underrepresented group). Other triggers capture information on the current or previous teaching period (e.g. the student is currently enrolled in

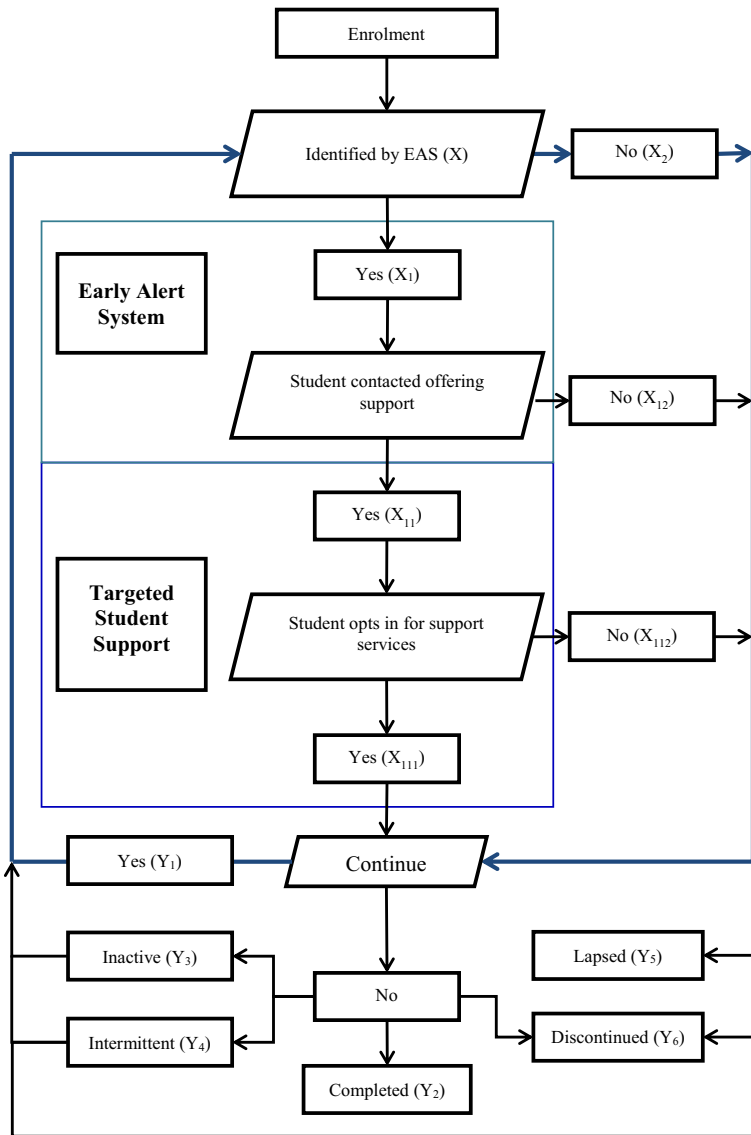


Fig. 1 Conceptual framework of the EAS Process at UNE

a high-risk unit). The most granular triggers analyse student log data, updating daily within the framework. Each trigger carries a positive or negative weighting which is added up each day, yielding a score which is assigned to the students. Students are then ranked based on their score attained, the highest 200 negative scores that day are considered as the most at risk of disengaging. These 200 at-risk students form the short list (X_1) are then passed to the student support team for further action. The sequence of events after identification relates to both if the student was contacted, and if so, whether the student chose to have tailored support. A key aspect of the system is the student’s decision to opt out of support, affording students self-determination in the support process after identification.

Method of analysis

Model selection

We use the survival analysis approach to examine the relationship between student retention and the EAS, among other contextual variables. Several studies have previously employed this approach to analyse student retention over time (DesJardins 2003; DesJardins et al. 1999; DesJardins et al. 2002; Ishitani 2006; Ishitani and DesJardins 2003). These studies analysed the factors associated with retention in different tertiary and temporal settings. Survival analysis allows appropriate treatment of variables in a temporal context where right censoring is an issue. For this study, survival is defined as a student being retained, with the term ‘failure event’ referring to a student discontinuing their studies through either events Y_5 or Y_6 in Fig. 1.

Survival analysis can be conducted using both parametric and non-parametric models. For the purposes of this study, the Cox proportional hazards model was chosen, on the basis that it is a non-parametric approach which imposes no assumptions on the underlying distribution of variables. Essentially, it estimates the hazard function $\lambda(t)$, which describes the level of risk associated with a defined failure event. Its main focus is in assessing ‘the relation between the distribution of failure time and $[x]$ ’ (Cox 1972, p. 189), where x captures the array of variables that can affect the hazard function. The Cox proportional hazards model extends to dealing with time-varying covariates that allows the estimated hazards associated with each explanatory variable (β) to change over time.

Censoring is another issue associated with survival analysis. In our study, this is where students discontinue their studies after the end of the period of analysis. The student would appear enrolled at the end of the data capture period; however, this does not represent the final outcome for the student. This is not a significant issue if the time period captured by the data set is sufficient. It is expected that within the period covered in the survival analysis, most students would discontinue to do so before the end of the 3-year data capture period, meaning that any error associated with censoring is minimal. Overall, the Cox proportional hazards model remains a useful approach of analysing the variables that affect student retention.

Key assumptions and tests

Students exit the model either by discontinuing their studies or through right censoring events. The main assumption of survival analysis is that proportional hazards remain constant over time. To correct the issues associated with time-varying covariates, the model allowed for interaction of these covariates over time. This isolates the effects of the independent variables to provide unbiased estimators. To identify which variables were time variant, the proportional hazards (PH) assumption test was run. If there was a significant relationship between any of the variables and time, the PH test indicates a p value less than 0.05.

Model specification

We estimate our empirical model using *demographic, institutional, student performance and workload* variables. The EAS variable is considered binary, where 1 represents a student that

was identified by the system at time t . This process is presented in Fig. 2. The model tests the question ‘is there a relationship between being identified by the EAS at a specific time t and the student’s survival rate’. Under the assumption that the system is functioning within its design parameters, there should be a statistically significant link between the EAS and the hazard ratio of students in this model.

Results

The estimates of the different hazard ratios are presented in Table 2. The results are divided according to different categories of variables, their respective hazard ratios and level of significance. In this table, the hazard ratio is defined as the chance of a variable affecting the decision of students to discontinue from their studies. The overall model results show that the model is significant at the 1% level. The PH test results show that, overall, the model has no significant correlation with time. This indicates that the model has not violated the proportional hazard assumptions associated with survival analysis.

Demographics

The estimated hazard ratios for demographic variables indicate that there is a significant relationship between *gender*, *age* and ATSI (Aboriginal Torres Strait Islander) status. The higher hazard ratio means that in any given week, females face a risk of discontinuing 7.6% higher than their male counterparts. *Age* is a significant variable expressed as a non-linear relationship by including the squared term. Using the estimates provided, it is possible to calculate the age at which students face the minimum hazard ratio. Up to the age of 51, the hazard ratio for students declines before it starts to rise again. This effect is important in the UNE context as the average age of students is around 29 years, suggesting that the current age profile of the university is helping maximise retention.

The ATSI variable is an important variable capturing people identifying as indigenous Australians. The result shows that students who are identified as ATSI have a hazard ratio around 14 to 16% lower than non-ATSI students. This significant result is likely due to additional on-campus support services provided to ATSI students. This finding shows that once cultural barriers around ATSI admission are removed, the ATSI students are more likely to continue and complete their studies in their given awards.

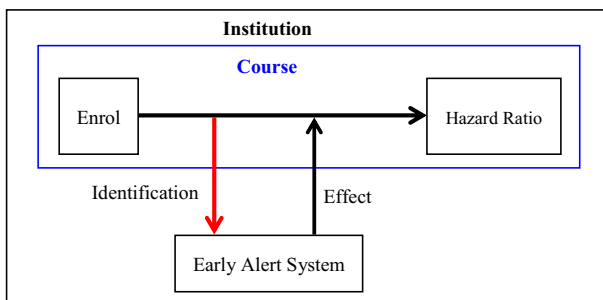


Fig. 2 Model on the relationship of EAS, identification and hazard ratio

Table 2 Estimates of hazard ratio for different categories of variables

Variable	Hazard ratio	Standard error
Demographics		
Gender	1.076 ^b	0.034
Age	0.969 ^a	0.007
Age ²	1.000 ^a	0.000 ^d
ATSI	0.837 ^b	0.070
Institutional		
Domestic fee	0.605	0.338
Domestic fee $\times t$	1.010	0.007
International fee	0.194 ^a	0.057
International fee $\times t^2$	1.000 ^a	0.000 ^d
International fee $\times t^3$	1.000 ^b	0.000 ^d
Prior studies	0.587 ^a	0.098
Prior studies $\times \ln(t)$	1.133 ^a	0.049
On-campus	1.175 ^a	0.055
On-campus $\times t$	0.910 ^a	0.013
Diploma	1.022	0.114
Advanced diploma	0.783 ^b	0.093
Advanced diploma $\times t$	1.003	0.002
Bachelors (graduate)	0.740 ^a	0.051
Bachelors (honours)	0.619 ^a	0.078
School		
School 1	1.127 ^c	0.078
School 2	1.272 ^a	0.091
School 2 $\times t^{-1}$	0.006 ^a	0.006
School 3	1.181 ^b	0.080
School 4	1.297 ^a	0.082
School 5	1.295 ^a	0.090
School 6	0.860 ^c	0.070
School 6 $\times t$	1.004 ^a	0.001
School 7	0.778	0.282
School 8	1.045	0.083
School 9	0.958	0.077
Student performance and workload		
Withdrawn	1.371 ^a	0.092
Withdrawn $\times t$	0.992 ^a	0.002
Withdrawn $\times t^2$	1.000 ^a	0.000 ^d
Withdrawn early	1.451 ^a	0.081
Withdrawn early $\times t$	0.992 ^a	0.002
Withdrawn early $\times t^2$	1.000 ^a	0.000 ^d
Fail incomplete	1.262 ^a	0.024
Fail incomplete $\times t$	1.000 ^c	0.000 ^d
Fail	1.141 ^a	0.020
Pass	0.847 ^a	0.012
Credit	0.891 ^a	0.013
Distinction	0.868 ^a	0.014
High distinction	0.874 ^a	0.018
Other	0.743 ^a	0.052
Inactive	45.447 ^a	5.592
Inactive $\times t$ (for $t \leq 16$)	0.136 ^b	0.106
Inactive $\times 1/\ln(t)$ (for $t > 16$)	0.084 ^a	0.016
Part-time	2.189 ^a	0.362
Part-time $\times \ln(t)$	0.976	0.054

Table 2 (continued)

Variable	Hazard ratio	Standard error
Early alert system		
Early alert system	2.067 ^a	0.214
Early alert system $\times t$	0.994 ^a	0.002
LR $\chi^2(50) = 7646.43$ (p value = 0.0000)		
PH test $\chi^2(50) = 39.12$ (p value = 0.8667)		

^aSignificant at the 1% level

^bSignificant at the 5% level

^cSignificant at the 10% level

^dRounded to three decimal places but not equal to zero

Institution

The institution variables capture five characteristics of student enrolment that relate to the institution, including *fee type*, *prior studies*, *study mode*, *course type* and *school of enrolment*. The base case for these five categories is a student using the Higher Education Loan Program (HELP) to pay their fees. The student has not previously studied and is studying off-campus using online resources. The student is studying a bachelor degree through normal admission paths in a school which offers professional qualifications. Some variables have hazard ratios of 1 and standard errors of 0 due to rounding. However, the significance of these variables is accurately presented and is an important inclusion in the models for completeness.

The variable *fee type* is categorical, using HELP students as the base category, which is the basis for comparison of domestic and international fee-paying students. The results show that there is no significant difference between the domestic fee-paying and HELP students, but there is significant difference between international fee-paying and HELP students. In passing note, the hazard ratio for international students is not constant over time.

We also examine the hazard ratio over time. Figure 3a shows the hazard ratio associated with an international student. The main coefficient estimate over time is plotted in blue, and this is bounded by the black dashed lines indicating the bounds of the 95% confidence interval over time. The red baseline shows that there is no significant difference or change in the hazard ratio of students. To effectively interpret the graph, statistically significant points occur upon the intersections of the black dashed confidence interval bounds and the baseline. The results show that international students start their studies with a significantly lower hazard ratio than HELP students. However, the hazard ratio increases over time until there is no significant difference around week 76, or one and a half years into their study. The hazard ratio continues to rise, peaking at 104 weeks of study. A key point is that there is no significant difference between international students and domestic HELP students between weeks 57 and 145. While the hazard ratio does increase, at no point are international students at a higher risk of discontinuing than HELP students. As such, while the highest risk of discontinuing occurs when students reach the 2-year mark, there is no cause for concern. This also indicates that if an EAS is to differentiate students based on international status, then it should reflect the decreased hazard faced by international students for the first year of study.

The variable *prior study* initially indicates a statistically significant relationship. However, upon calculation of the combined temporal effect, there is no statistically significant

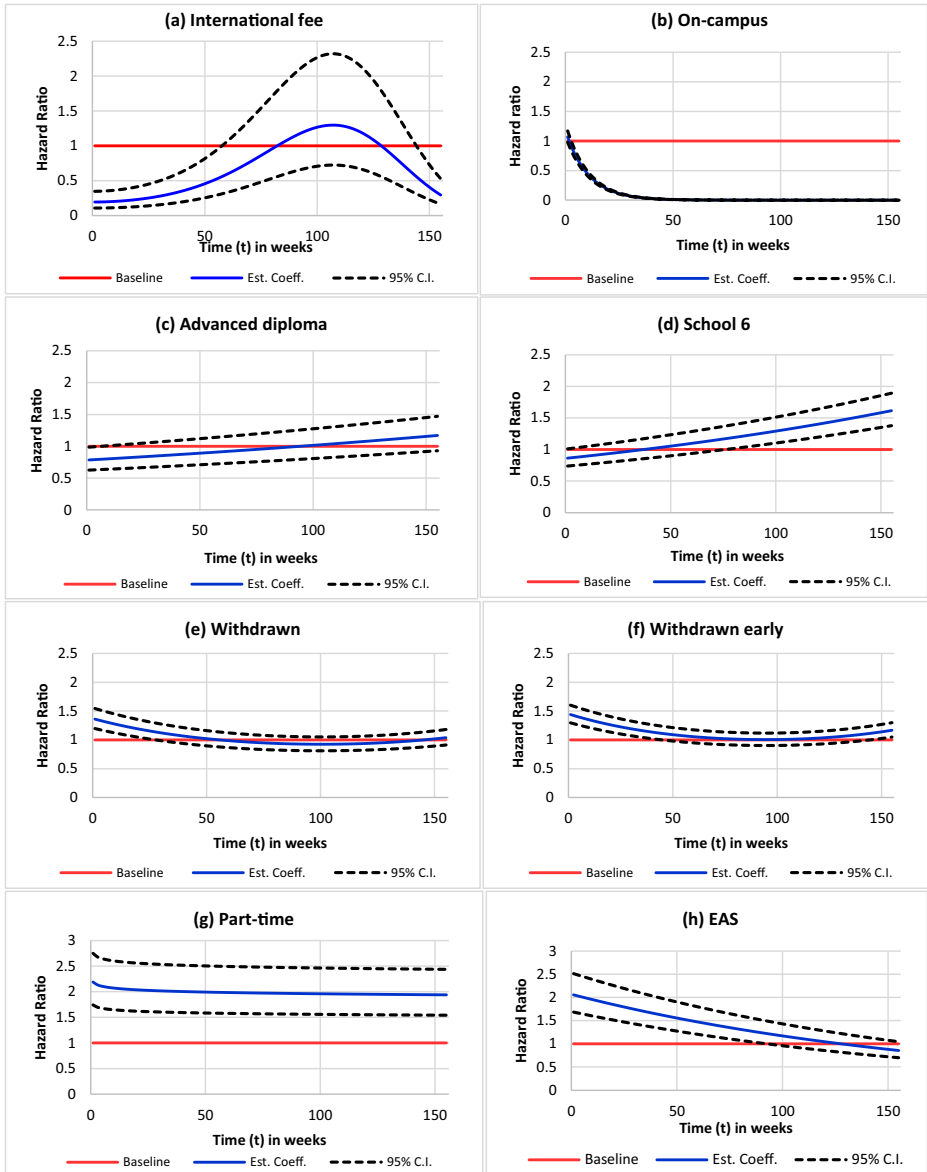


Fig. 3 Hazard ratios over time for selected variables. **a** International fee. **b** On-campus. **c** Advanced diploma. **d** School 6. **e** Withdrawn. **f** Withdrawn early. **g** Part-time. **h** EAS

relationship present. This result highlights technical issues surrounding time series analysis. This is important; theoretically, there should be some significant relationship. The absence of a relationship indicates several potential issues. More thought may be required on how the prior studies are measured as a variable and included in retention modelling, or the theoretical assumptions around the relationship between prior studies and student retention need to be revised.

The *on-campus* variable compares students who complete their studies attending classes in person at the institution versus off-campus online students. The estimated hazard ratio is presented in Fig. 3b. Initially, on-campus students have a higher hazard ratio than off-campus online students. However, the effect rapidly decreases after a few weeks, with on-campus students having a significantly lower hazard ratio after 3 to 5 weeks. This shows a stark divide in the risk for students in different modes of study.

Comparing degree types, there is no significant difference in the hazard ratios of diplomas and the base case of the bachelor degree. For *advanced diploma* courses, Fig. 3c shows an interaction effect over time, with a reduced hazard ratio than normal entry bachelor students at the start of the course at the 5% significance level. However, this significant difference is only for the first 5 weeks. A significant difference is found between students admitted through the University Admission Centre (UAC) process and *bachelor students* who are admitted through graduate entry or admitted for honours programmes. Students admitted in these latter cases have a significantly lower hazard ratio at the 1% significance level. Both estimates are in line with expectations that students in these categories are better prepared for university study.

With regard to *Schools*, the results indicate variation in the hazard ratios compared to the base school. Additionally, *schools 2* and *6* have temporal effects where the hazard ratio is time variant. *School 1* has a significantly higher hazard ratio at the 10% level from the base school. *School 3* is significantly higher at the 5% level, while *schools 4* and *5* are significantly higher at the 1% level. Although *schools 2* and *6* have time-varying dimensions, the combined hazard ratio in the former is not significantly different from the baseline value of 1. In the case of *school 6*, Fig. 3d shows that the hazard ratio increases linearly over time. Initially, students in this school have a lower estimated hazard ratio than the base school, but this is not statistically different, as indicated by the confidence interval bounds. However, the upward trend indicates that by the 80th week of enrolment, there is a significant difference in the hazard ratios from the base school, indicating that the longer the student is enrolled in *school 6*, the higher the risk that student will discontinue.

Student performance and workload

The two learning environment variables used in this model include final grades of the student and the level of workload undertaken by the student. The grades can be divided into three classes, namely, negative grades (*withdrawn*, *withdrawn early*, *fail incomplete* and *fail*) which do not contribute to a student's progression in the course, positive grades (*pass*, *credit*, *distinction* and *high distinction*) which indicate a level of competency in the unit, and *other* administrative grades which capture more complex student situations such as special extension of time or special examinations.

The estimated coefficients show the marginal effects; thus, if a student receives both a credit and a distinction upon completing two units of study, the overall changes to the base hazard ratio would be calculated by adding the logarithm of the two coefficients. Furthermore, the effect of attaining a *withdrawn*, *withdrawn early* and *fail incomplete* changes over time. For the *withdrawn* grade, Fig. 3e shows the non-linear relationship of the hazard ratio over time, with initial effects indicating students who withdraw during their first year having a significantly higher hazard ratio. However, after 26 weeks, the results show no significant effect resulting from withdrawing. This is logical given that students who withdraw during their first teaching period are likely to be struggling with study, while students withdrawing later may indicate gaming of the system.

The *withdrawn early* grade occurs when students withdrew before the financial census date. As such, the student is not negatively affected financially or academically with respect to grade point average, and the relationship was found to be non-linear (see Fig. 3f). However, receiving a *withdrawn early* grade increases the hazard ratio for around the first 42 weeks of enrolment. After this point, a *withdrawn early* grade is not indicative of an increased discontinuation risk again until around week 147. This shows that students who withdraw early from units within their course do exhibit a higher hazard of discontinuing, especially for the first year and a half of study. The difference between *withdrawn* and *withdrawn early* grades is that, for the latter, the financial cost of the unit is not passed onto the student. Therefore, the difference between the hazard ratios for the two grades may indicate the magnitude of the effect of fee payments on retention.

The *fail incomplete* grade is also time-dependent with hazard ratios. The results show that receiving a *fail incomplete* earlier in the enrolment is associated with a higher hazard ratio. While the effect decreases over time, it remains a significant variable over the 3 years of data captured in the model. *Failure* grades are far more consistent over time. For each *failure* grade, the students hazard ratio increases by 14% which is constant over time.

Positive grade outcomes of *pass*, *credit*, *distinction* and *high distinction* all have a positive effect in reducing the hazard ratio of the student, with the magnitude of the effect relatively constant. This indicates the positive effect of progressing with a course. Interestingly, students who receive a grade in the *other* category also have a reduced hazard ratio. Most of these grades require a degree of administrative interaction, which could assist with academic integration. Therefore, grades that progress students in their degree have a positive effect on reducing risk associated with discontinuation.

Workload is a statistically significant variable that affects the hazard ratio. Students can have three possible levels of workload: *full-time study*, defined as the student enrolls in three or more units of study per teaching period; *part-time study*, where the student enrolls in fewer than three units per teaching period; and *inactive*, where a student undertakes no units of work but is enrolled in a course. The base case is the full-time student. Comparing full-time to inactive students captures the effects associated with students who take a break from their studies and return at a later date. The results show that this variable is highly correlated to two variables capturing interactions over time. The complex time-varying aspects associated with inactivity

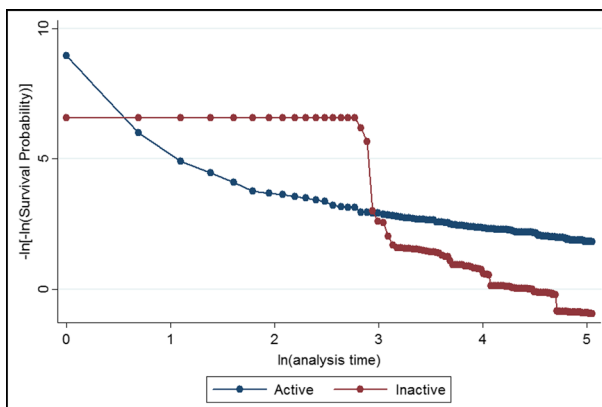


Fig. 4 Survival probability of inactive students

are shown in a Kaplan-Meier graph in Fig. 4, comparing the probability of survival for inactive and non-inactive students.

If the proportional hazards assumption is upheld in Fig. 4, the two lines should be parallel. Clearly, this is not the case, with a linear function up to week 16, after which the function takes on an inverse logarithmic shape. This indicates a discontinuous function over time, so two equations for inactivity are required for the model to function correctly. The first equation captures the linear effect over the first 16 weeks. The second equation captures the inverse logarithmic function from week 17 onwards. The estimated hazard ratio for inactivity is plotted over time incorporating both functions to create Fig. 5.

Compared to the other estimates presented in Fig. 3, the magnitude of the effect of inactivity on hazard ratios is very large. Initially, the model indicates that inactivity at the start of enrolment increases the hazard ratio significantly. This decreases to near 0 levels until week 18, after which the hazard ratio increases dramatically. The ditch in the hazard ratios can be explained by the limited number of inactive observations during these first initial weeks of teaching. After the first teaching period has been completed, a dramatic increase in the hazard ratios commences. It shows that inactivity is the largest contributing factor to a student’s hazard ratio after 18 weeks of enrolment. As such, this is an essential variable that needs to be captured as part of an EAS.

The other mode of study captured by this study is part-time study. Visually, the hazard ratios are presented in Fig. 3g. Part-time students initially have a significantly higher hazard ratio, which decreases over time. However, after 3 years, the hazard ratio is still significantly greater than 1. This indicates a strong disparity in the risk of discontinuing between part-time and full-time students. The results show that decreasing the workload of the student increases the hazard ratio and risk associated with discontinuing. This acts as a strong indication that the students’ mode of study reflects their level of commitment to their course and remaining enrolled. The results indicate that full-time enrolment minimises the risk associated with discontinuing.

Early alert system

Having taken into account other explanatory variables, the model is able to identify the relationship between the EAS and the risk of discontinuing. Figure 3h presents the computation of the joint interaction among individual coefficients with 1% significance level overtime.

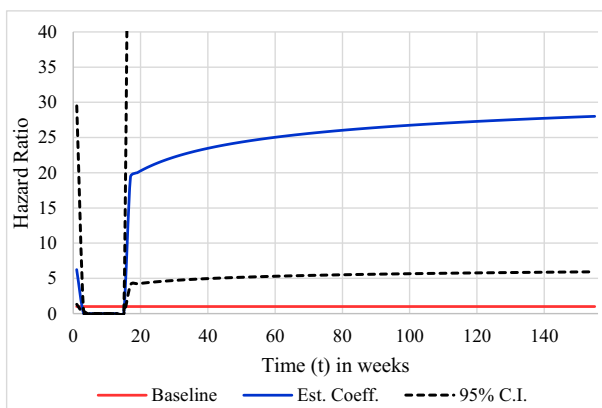


Fig. 5 Inactivity hazard ratio over time

The model shows that students identified by the EAS are most at risk of discontinuing when first enrolled. This provides strong evidence that those identified by the system have a higher risk of discontinuing compared to those not identified by the system. Additionally, the effect is most pronounced at the start of enrolment and gradually decreases over time. Figure 3h shows that at around 90 weeks, students identified by the EAS have no significant difference in the hazard ratio from students not identified by the EAS. This indicates the EAS may struggle to differentiate and identify students at risk of discontinuing beyond this point in time. Two reasons go to explaining this: (1) The number of students who discontinue in the later years of study is significantly less; (2) this could be a crowding out effect of the algorithm, where the daily list of the top 200 students is populated by students in earlier stages of enrolment. As such, it may be harder for EAS systems to identify these students who choose to discontinue, despite having significant course progression.

Summary and conclusion

We used a survival analysis approach to examine the relationship between student retention and an EAS, along with key *demographic, institutional, student performance* and *workload* variables. We used the hazard ratios to indicate the risk of students discontinuing in enrolment. Significant findings show that *gender, age* and *ATSI* status are important variables to include in studying student retention. In the institutional dimensions, any EAS should account for risk changing over time. This is highlighted in the case of *international fee-paying* students, where the hazard ratio has a rapid increase close to completion of 2-year courses. This also supports the use of survival analysis to capture detailed effects over time. When comparing course types, *advanced diploma* students also have a significant difference from normal *bachelor entry* students. The hazard ratio is not constant and increases the longer the student is enrolled. Students who enter through *graduate entry* or directly into *honours* programmes have lower hazard ratios than bachelor students admitted through traditional entry path. The variation between *schools* shows that there is scope to incorporate school-specific effects within the EAS. Specifically, *school 6* needs further analysis to establish the significant issues causing the increase in the hazard ratio over time. The between-school differences also represent a level of necessary customisation that EAS should undertake when deployed at an institution level.

In considering student performance, negative grade outcomes need to be adequately accounted for. Given that *withdrawn early* grades occur within the first few weeks of a teaching period, before the financial census date, this should be a major predictor for timely identification of a student who is intending to discontinue their studies. However, timing is also important, with the observed effect of *withdrawn early* grades only occurring within the first year of study. Furthermore, the model shows that the student performance can be treated cumulatively as a sum of the estimated coefficients for grades. The results show that a student who attained three passes and a fail in a teaching period would have a reduced hazard ratio overall. Inactivity of students also needs to be factored into the EAS algorithm. Periods of inactivity, regardless of whether the institution is informed of the student's intent to take leave from studying, indicate a significantly higher hazard ratio across most time periods. In terms of its magnitude, inactivity contributes the largest increase in the estimated hazard ratio. Therefore, the designers of an EAS must capture inactivity as a major predictor of discontinuation.

With respect to an EAS, results indicate that the system is identifying students at risk of discontinuing their studies. This effect decreases over time, however, indicating that the EAS

algorithm is not able to identify students at risk of discontinuing beyond week 90 of enrolment. Given that the EAS design was to focus on identifying students at risk of disengagement, and not discontinuation, this result is not a major concern for the EAS algorithm. It demonstrates a link between the system and the risk of discontinuing. Furthermore, it shows that the EAS identification process is functioning correctly during the critical first year of student enrolment. This provides strong empirical support for the utilisation of EAS to assist student support teams in providing targeted support to students in need of assistance.

This study presents important significant findings regarding the relationship between EAS and student retention. This study makes important contributions to not only both the understanding of the complex relationship between EAS and student retention but also how survival analysis can be used to create a detailed understanding of these factors, including how the EAS is performing in identifying students. Several lines of future research can be identified from this study. This includes expanding the analysis to include other variables that relate to student retention not captured in this study, such as more detailed information family responsibilities and the income arrangements beyond just fee categories. From the institutional perspective, incorporating other intervention vectors, such as meetings with faculty members and academic mentors, could provide a more detailed understanding of this complex environment. Future work can also focus on analysing how different representations of the EAS variable can affect both the model estimates and interpretations of how the system is functioning. This would include using an enduring EAS effect, which would look at the hazard ratio after the first instance of being identified. Another analysis would be to separate students into the overall 'identified' and 'not identified' categories, and test for significant differences between the two groups of students. This could also be expanded to an interaction effects model. A second area of analysis is to directly regress the triggers of the EAS algorithm to the hazard ratio. This would help to identify which underlying triggers have a direct statistical connection to the survival rate and identify which aspects of the algorithm may be most useful in developing a retention-specific EAS.

Acknowledgements The authors acknowledge the support of the staff at the case study institution in making an expansive data set available for analysis. We also wish to acknowledge the three anonymous reviewers and the editor for their helpful comments and suggestions. The usual caveat applies.

References

- Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: using learning analytics to increase student success. In Proceedings of the Second International Conference on Learning Analytics and Knowledge. ACM. 267–270.
- Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: the incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies.*, 68(4), 223–241.
- Baker, R. S., Lindrum, D., Lindrum, M. J., and Perkowski, D. (2015). *Analyzing early at-risk factors in higher education e-learning courses*. Retrieved 12/09/2015 from <http://www.columbia.edu/~rsb2162/2015paper41.pdf>
- Caufield, M. (2013a). *A simple, less mathematical way to understand the course signals issue*. Hapgood. Retrieved 10/07/2015, from <http://hapgood.us/2013/09/26/a-simple-less-mathematical-way-to-understand-the-course-signals-issue/>
- Caufield, M. (2013b). *Why the course signals math does not add up*. Hapgood. Retrieved 08/12/2014 from <http://hapgood.us/2013/09/26/why-the-course-signals-math-does-not-add-up/>
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 187–220.

- DesJardins, S. L. (2003). Event history methods: conceptual issues and an application to student departure from college. *Higher Education: Handbook of Theory and Research*, Springer. 421–471.
- DesJardins, S. L., Ahlburg, D. A., & McCall, B. P. (1999). An event history model of student departure. *Economics of Education Review*, 18(3), 375–390.
- DesJardins, S. L., Ahlburg, D. A., & McCall, B. P. (2002). A temporal investigation of factors related to timely degree completion. *The Journal of Higher Education*, 73(5), 555–581 Retrieved from <http://www.jstor.org/stable/1558433>.
- Ferguson, R., Clow, D., Macfadyen, L., Tynan, B., Dawson, S., & Alexander, S. (2015). Setting learning analytics in context: overcoming the barriers to large-scale adoption. *Journal of Learning Analytics*, 1(3), 120–144.
- Harrison, S., Villano, R., Lynch, G., & Chen, G. (2016). Measuring financial implications of an early alert system. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 241–248). ACM.
- Ishitani, T. T. (2006). Studying attrition and degree completion behavior among first-generation college students in the United States. *The Journal of Higher Education*, 77(5), 861–885 Retrieved from <http://www.jstor.org/stable/3838790>.
- Ishitani, T. T., & DesJardins, S. L. (2003). A longitudinal investigation of dropout from college in the United States. *Journal of College Student Retention*, 4(2), 173–201.
- Jayaprakash, S. M., Moody, E. W., Lauria, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: an open source analytics initiative. *Journal of Learning Analytics*, 1(1), 6–47.
- Jones, K. M., Thomson, J., and Arnold, K. (2014). *Questions of data ownership on-campus*. Educause Review. Accessible online at <http://www.educause.edu/ero/article/questions-data-ownership-campus>.
- Khamis, C., & Kiernan, F. (2013). *Track and connect: a tailored individual support program for at-risk students at the University of Sydney*. Sydney: University of Sydney Retrieved from http://fyhe.com.au/past_papers/papers13/6F.pdf.
- Krumm, A. E., Waddington, R. J., Teasley, S. D., and Lonn, S. (2014). A learning management system-based early warning system for academic advising in undergraduate engineering. In *Learning analytics*. Springer New York. 103–119.
- Lawson, C., Beer, C., Rossi, D., Moore, T., & Fleming, J. (2016). Identification of ‘at risk’ students using learning analytics: the ethical dilemmas of intervention strategies in a higher education institution. *Educational Technology Research and Development*, 64(5), 957–968.
- Leece, R., and Cooper, J. (2011) *Proactively managing student wellbeing at UNE*. AAIR data warehouse, load management, DEEWR SIG forum, Australian National University, published by Kearns, C. Accessible online at <https://prezi.com/m5qui5cptvay/copy-of-unes-automated-student-wellness-engine/>
- Marrington, A. D., Nelson, K. J., and Clarke, J. A. (2010). *An economic case for systematic student monitoring and intervention in the first year in higher education*. Paper presented at the proceedings of 13th Pacific rim first year in higher education conference, Adelaide. Retrieved from <http://eprints.qut.edu.au/33231/1/c33231.pdf>
- Nelson, K. J., and Creagh, T. A. (2013). *A good practice guide: safeguarding student learning engagement*. Brisbane: Queensland University of Technology. Retrieved 08/06/2015 from http://eprints.qut.edu.au/59189/1/LTU_Good-practice-guide_eBook_20130320.pdf
- Prinsloo, P., & Slade, S. (2016). Student vulnerability, agency and learning analytics: an exploration. *Journal of Learning Analytics*, 3(1), 159–182.
- Purdue University Information Technology (2013) *Course signals* Retrieved 11/02/2015 from <http://www.itap.purdue.edu/learning/tools/signals/>
- Romero-Zaldivar, V.-A., Pardo, A., Burgos, D., & Kloos, C. D. (2012). Monitoring student progress using virtual appliances: a case study. *Computers and Education*, 58(4), 1058–1067.
- Siemens, G., Dawson, S., and Lynch, G. (2013). Improving the quality and productivity of the higher education sector: policy and strategy for systems-level deployment of learning analytics. Sydney. Office for learning and teaching. Retrieved from http://www.olt.gov.au/system/files/resources/SoLAR_Report_2014.pdf
- Straumsheim, C. (2013). Mixed signals. Inside higher Ed. Retrieved from <http://www.insidehighered.com/news/2013/11/06/researchers-cast-doubt-about-early-warning-systems-effect-retention>
- Tynan, B., and Buckingham Shum, S. (2013). *Designing systematic learning analytics at the Open University*. SoLAR Open course, strategy and policy for systematic Learning Retrieved from <http://www.slideshare.net/sbs/designing-systemic-learning-analytics-at-the-open-university>