Microeconometric Approaches in Exploring the Relationships Between Early Alert Systems and Student Retention: A Case Study of a Regionally Based University in Australia

Scott Harrison[1], Renato Villano[2], Grace Lynch[3], George Chen[4]

Abstract
Early alert systems (EAS) are an important technological tool to help manage and improve student retention. Data spanning 16,091 students over 156 weeks was collected from a regionally based university in Australia to explore various microeconometric approaches that establish links between EAS and student retention outcomes. Controlling for numerous confounding variables, significant relationships between the EAS and student retention were identified. Capturing dynamic relationships between the explanatory variables and the hazard of discontinuing provides new insight into understanding student retention factors. We concluded that survival models are the best methods of understanding student retention when temporal data is available.

Notes for Practice

- We use predictive analytics to investigate, analyse and understand the demographic, institution, learning environment and temporal variables affecting undergraduate student retention rates.

- Early alert systems (EAS) can form an important mechanism to address student retention, but the EAS must also be designed with retention in mind. The wellness engine was and remains primarily focused on identification for support needs, not on student retention.

- Temporal models such as survival analysis, is required to capture the complexities of changes over time. Treatment effects models serve an important function in decomposing effects when randomized controlled trials is not an option, as is frequent with student retention and the allocation of support resources.

Corresponding author [1]Email: harrison@dipf.de Address: DIPF, Leibniz Institute for Research and Information in Education, Rostocker Straße 6, 60323 Frankfurt am Main, Germany and UNE Business School, University of New England, Armidale, NSW, 2351, Australia. ORCID ID: https://orcid.org/0000-0001-6712-7784
[2]Email: rvillan2@une.edu.au Address: UNE Business School, University of New England, Armidale, NSW, 2351, Australia. ORCID ID: https://orcid.org/0000-0003-2581-6623
[3]Email: glynch14@gmail.com Address: UNE Business School, University of New England, NSW 2351, Australia. ORCID ID: https://orcid.org/0000-0003-1508-6815
[4]Email: gchen2@une.edu.au Address: UNE Business School, University of New England, Armidale, NSW, 2351, Australia. ORCID ID: https://orcid.org/0000-0002-8519-9340

1. Introduction

Technological approaches to maximizing student retention are becoming increasingly common in many educational domains. Importantly, technology-driven programs can help minimize the number of students discontinuing, while assisting students toward other goals, such as improving learning outcomes and welfare. One such approach is implementing an early alert system (EAS), software incorporating a set of predetermined risk factors for analyzing data and identifying students at risk. Implementing this type of learning analytics solution holds great promise. Indeed, such technologically innovative processes are currently used by many universities in various forms.

Despite their rising popularity, EASs' effectiveness on student retention in the higher-education sector remains a highly contentious issue. To expand the current understanding of the EAS–retention nexus, this study applies various microeconometric methods to the data collected from a regionally based Australian university to identify how EASs and student retention are linked. Specifically, three key research questions form the focus of this study: (1) What variables affect undergraduate students' retention rate? (2) How does an EAS influence the retention of students at the university? (3) Which microeconometric method represents the best approach for identifying the relationship between EASs and undergraduate outcomes?

To answer these questions, we review previous studies on student retention, technological adaption, and EAS, and outline the case study institution's background. In the Methods section, we introduce our analytical framework for estimating the EAS–retention nexus and the various modelling approaches to addressing the research questions. The Results section details the mixed-method results, covering the control variables and the main EAS effects. Finally, the Discussion section contextualizes the findings within the learning analytics domain and suggests directions for future research.

1.1. Previous Works

Development of literature in the field of student retention has been a keen research interest since the 1920s, and especially since the 1960s, meaning that "the literature on student retention is voluminous and arguably capricious" (Simons, 2011, p. 13). This wide range of research occurred due to the wide variety of academic disciplines contributing to research in student retention, including psychology, education, and social science. Several foundational studies by Tinto (1975), Bean (1980), Astin (1984), and Bean and Metzner (1985) developed frameworks and tried to define models of student retention that could be universal; summarizing the contribution of the theories, the models incorporated several common themes, none more important than the role of institutional and social involvement in student outcomes, changing the discourse on student retention to acknowledge that universities could influence and affect student outcomes. The theories laid the foundation for institutions to take proactive steps to determine how to improve student retention. These theories were the precursors to the current body of literature developed in learning analytics, in particular for EASs that aim to affect retention.

With a strong statistical focus, empirical studies on learning analytics are of increasing importance in understanding student retention. The implementation of IT systems has enabled universities to collect massive datasets on students and their interactions with the learning environment, both of which influence decision-making within the higher-education sector (Siemens, Dawson, & Lynch, 2013; Parnell, Jones, Wesaw, & Brooks 2018). One aspect of learning analytics has been to use data to identify students in need of targeted support (Aguilar, Lonn, & Teasley, 2014; Arnold, Tanes, & King 2010; Jayaprakash, Moody, Lauría, Regan, & Baron, 2014). Targeted support systems require significant university resources in development, deployment, and ongoing operations. However, from the perspective of universities and students, the effect of student support is generally unquantified. Indeed, Ferguson and Clow (2017, p. 8) analyze the evidence supporting learning analytics and conclude that "the nature of the topic area makes it hard to carry out rigorous quantitative research." In general, a fundamental problem for universities is justifying these complex programs and determining what, if any, effect they have had on student retention.

Several quantitative studies on student retention have been conducted to date, measuring variables that affect student learning outcomes objectively. Previous studies include linear regressions (Harrison, Villano, Lynch, & Chen, 2014), probit and logit models (Lin, Yu, & Chen, 2012; Jones-White, Radcliffe, Huesman, & Kellogg, 2010; Singell & Waddell, 2010), two-stage sequential decision models (Stratton, O'Toole, & Wetzel, 2007), treatment effects modelling (Harrison, Villano, Lynch, & Chen, 2016), and survival models (Villano, Harrison, Lynch & Chen, 2018; DesJardins, Ahlburg, & McCall, 2002; Ishitani 2003, 2006; Ishitani & DesJardins, 2003). Among these competing models, the survival models have the additional advantage of capturing temporal effects in student retention data. The relationship between independent and dependent variables can change over the time span of analysis. To allow an objective valuation of the data using cross-sectional and panel data, we have chosen a mixed-method approach to search for converging evidence of how EASs are linked to student retention outcomes.

2. Methods

When applying the mixed-method approach, it is important to contextualize the environment for data collection. With the focus on EASs and student retention, various EASs have been implemented across Australian universities (Marrington, Nelson, & Clarke, 2010; Nelson & Creagh, 2013). The case university is one of 40 in Australia and part of the Rural Universities Network. In 2013, its enrolment reached 22,389 students, of which 78.9% were off-campus students. This share of off-campus student cohort was significantly higher than national averages (Department of Education, 2014), making this university unique in the higher-education sector. Furthermore, it has created a strong focus on supporting students from low socioeconomic

backgrounds through distance education. In early 2011, an institution-wide EAS was designed and deployed, which provided daily reports to the student-support group on students at risk of disengagement. Figure 1 is a flow chart describing the EAS decision-making process for the students and the support team, which was initially described by Villano et al. (2018) to identify the critical steps in the decision-making process.

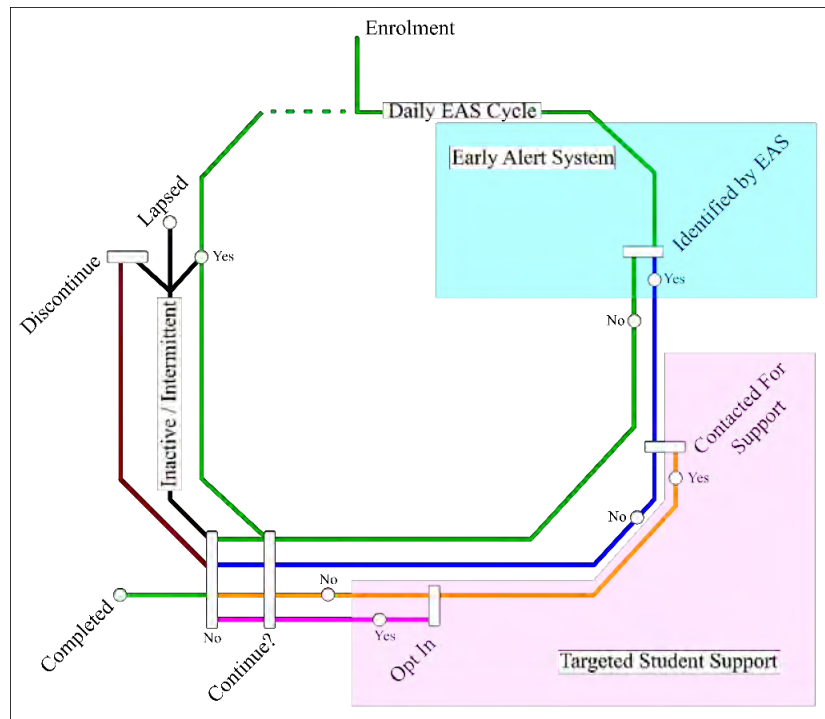**Figure 1.** Early alert decision cycle (Adapted from Villano et al., 2018).

Figure 1 describes a significant opening limitation on this dataset because data was only available for the initial identification process, not for whether students accessed the targeted student support. However, this still provides a rich dataset that forms the basis of our study.
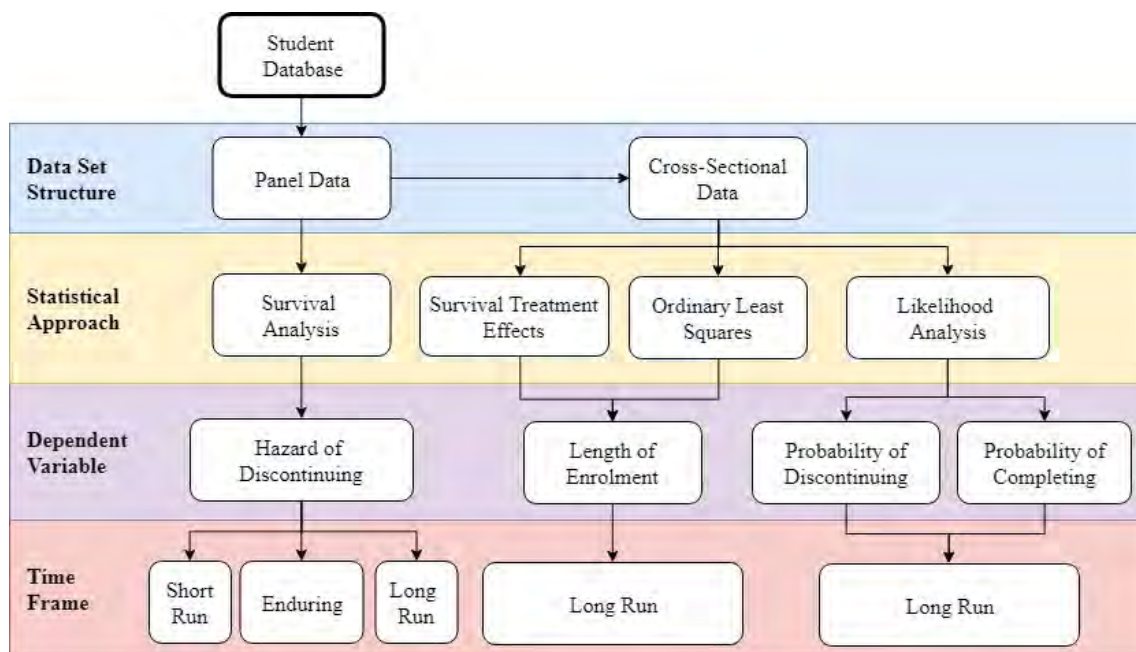

Figure 2. Analytical framework.

We employed a mixture of model specifications and microeconometric analyses to estimate the EAS–student retention nexus in the case study institution. Figure 2 outlines each statistical approach's attributes, including data, dependent variable, and the relevant sample period.

Our study advances previously unpublished work on the survival treatment effects (STE) model (Harrison, 2016). It introduces two additional model specifications, which capture nuanced differences in how the EAS is represented. In Figure 2, the survival analysis models estimated in Villano et al. (2018) focus on the short-run effect in the time frame, that is, the effect of each time a student was identified by the EAS. The key findings from Villano et al. (2018) validate the EAS in linking each identification by the EAS to an increased chance of discontinuing. This analysis expands on this by including enduring and long-run survival models to see how results converge or diverge as a result of the EAS representation. The enduring model differentiates students from the time of first being identified by the EAS, and, as such, once the EAS identifies a student, they are considered "at risk" from that moment on. In the long-run model, students are deemed "at risk" if they are ever identified, across all periods. Doing so means that the model splits students into two categories independent of time: either identified or not identified by the EAS.

Additionally, we revise the estimates for the previously estimated ordinary least squares (OLS) model (Harrison et al., 2014) and maximum-likelihood models (Harrison, Villano, Lynch, & Chen, 2015). The previous work by Harrison et al. (2014) showed that by using linear regression, students' length of enrolment was linked to identification by the EAS, with identified students being enrolled for longer. The previous maximum-likelihood models (Harrison et al., 2015) showed that students identified by the EAS had a higher probability of discontinuing than those who were not identified by the EAS. Both models also included various covariates to help identify other potential relationships. In this study, these models are re-estimated by aligning the measurement models of the panel data analysis to contain the same underlying dataset and variables as used in the survival analysis. Doing this creates models that should be comparable at an inferential level.

Our dataset contains student-level attributes, including demographic, institutional, student performance, and study load variables. We include the EAS as a variable to assess its effectiveness in promoting student retention. Linking these variables together, we derive seven models from the following general specification:

$$Ym = \beta_0 + \beta_p D_p + \beta_q I_q + \beta_r P_r + \beta_s L_s + \beta_i EAS + \varepsilon \quad (1)$$

where Y is the statistical approach-based outcome variable, with m representing the matrices for the demographic (D), institution (I), performance (P), and study load (L) variables and their corresponding coefficients, β_p, β_q, β_r, and β_s, and EAS represents the EAS variable specified for each statistical approach, with its coefficient β_i reflecting the effect of the EAS identification. In terms of student attributes, the demographic (D) variable accounts for the student's background at the time of enrolment. The institutional (I) variable reflects the interactions between the student and the institution, including fee category, course type, and school of enrolment. The performance (P) variable uses the student's academic record to reflect varying grade outcomes in the learning environment. The study load (L) variable adjusts for the student's rate of progression in the chosen course of study. Intuitively, since these variables also capture factors affecting student retention, we control them to ascertain the true effect of the EAS on retention. Finally, β_0 is the constant term and ε is the stochastic error term.

Equation (1) serves as the basis for comparing the approaches taken to understand how EAS affects student retention. Table 1 summarizes the interpretation of the dependent variables concerning the definition of EAS under each approach.

Table 1. Summary of dependent variable interpretations

Model type	Time frame	Dependent variable interpretation
Survival	Short run	The hazard associated with a student discontinuing their studies at time period t, *if* they were identified by the EAS *in any given week*
Survival	Enduring	The hazard associated with a student discontinuing their studies at time period t, *once* they were identified by the EAS
Survival	Long run	The hazard associated with a student discontinuing their studies at time period t, *given* they were identified by the EAS
STE	Long run	The length of time a student is enrolled, given they were identified by the EAS
OLS	Long run	The length of time a student is enrolled, given they were identified by the EAS
Likelihood — discontinuing	Long run	The probability of a student discontinuing their studies, given they were identified by the EAS
Likelihood — completing	Long run	The probability of a student completing their studies, given they were identified by the EAS

In the cross-sectional analysis, two model variants are executed for the STE, OLS, and maximum-likelihood models. Specifically, the first variant specifies the EAS as a binary identified/not-identified variable, while the second classifies students based on the number of identifications. Since the estimated models produce near-identical estimates for all control variables, the results indicate the effects only for the binary EAS approach.

From the survival modelling, the different sample period is constructed using a binary variable. For the short-run model, the EAS variable assumes a value of one in any given week when the EAS identified the student. The enduring model assigns a value of one every week after the EAS first identified the student. The long-run model splits the student cohort into identified/not-identified groups for all periods across the entire analysis, regardless of when the student was identified as being at-risk. Table 2 is an example of the dataset structure for a sample student identified as at risk in week 2 of the study. Notice that in the short-run model, a value of one is assigned to this student's EAS variable when they were identified in that week. If it is the first time the EAS identifies this student, it also triggers the enduring model. For the long-run model, this student will always be considered at risk and assigned a value of one, regardless of when the identification occurred.

Table 2. Example EAS variable representation

Time frame	Short run	Enduring	Long run
Week 1	0	0	1
Week 2	0	0	1
Week 3	1	1	1
Week 4	0	1	1
⋮	⋮	⋮	⋮
Week n	0 or 1	1	1

In the survival analysis approach, we encounter two challenges, namely, in interpreting the estimated dependent variable, that is, the hazard ratio, and in capturing time-varying components (TVCs). The hazard ratio is an unobserved variable capturing both the probability and the timing of an event occurring at a given point in time. In this context, it can be seen as the risk of a student discontinuing under the effect of an independent variable at a given period (t). In turn, it introduces TVCs, where some effects relating to student retention may be constant throughout a student's enrolment. However, a more likely scenario is that the variables defined in the general model will vary over time. We test the survival models using the proportional hazards assumption tests (presented in the Appendix) to determine whether all TVCs are captured in the model. Reporting of independent variables with identified TVCs includes time (t) as part of the estimations.

Our survival models improve on Villano et al. (2018) by incorporating the study load (L) variable. Unlike Villano et al. (2018), who used a categorical variable to classify a student's status in each period as full-time, part-time, or inactive, we use the actual number of units a student is enrolled in during any given week. While the former approach is easier to manage, inactive students' effects are more difficult to interpret, as these students typically discontinue studies during teaching breaks where no units of study are taken, further confounding the variable. Instead, we overcome this complication by considering the actual number of units enrolled, which is a more accurate representation of the study load and the associated progress a student is making on their work.

In summary, we estimate the EAS–student retention nexus by the mixed-method approach. The results are nuanced to the context of each model, and, as such, we must interpret the empirical results with caution to avoid erroneous comparison.

3. Results and Discussion

3.1. Control Variables

In theory, controlling variables accounts for unobserved, confounding effects that enable the isolation of the relationship between the EAS and student retention in equation (1). This section reports the results from the enduring and long-run survival models, revises estimates for the OLS and likelihood models, and compares these results to Villano et al. (2018).

3.1.1. Demographic Variables

Table 3 compares the estimates of three key demographic (D) variables in equation (1), namely, gender, age, and Aboriginal and Torres Strait Islander (ATSI) status. Although gender should exert no discernable effects on student outcomes in theory, numerous studies have shown a close relationship between them (Tinto, 2006, p. 3). We also observe this apparent contradiction in our study. For example, our OLS estimate shows that gender is independent of the students' length of enrolment. However, our maximum-likelihood estimates suggest that, compared to their male colleagues, not only are female

students 10.4% more likely to discontinue, but they are 20.3% more likely to complete their studies. Such gender differences in discontinuing and completing are reflected in our survival models, with female students having a 6.1% higher hazard of discontinuing their studies. Our mixed results highlight the complexity of gender classification in the learning environment. Within the suite of university support options, there may be a need to consider gender-specific services.

Table 3. Estimates of the demographic variables

Model	Survival: short run[d]		Survival: enduring		Survival: long run		OLS: weeks enrolled		Likelihood of discontinuing		Likelihood of completing	
	HR	SE	HR	SE	HR	SE	b	SE	RRR	SE	RRR	SE
Gender (f = 1)	1.06[c]	0.03	1.06[c]	0.03	1.06[c]	0.03	0.32	0.33	1.10[b]	0.05	1.20[b]	0.11
Age	0.96[a]	0.01	0.96[a]	0.01	0.96[a]	0.01	0.23[b]	0.09	0.96[a]	0.01	1.06[a]	0.02
Age^2	1.00[a]	0.00	1.00[a]	0.00	1.00[a]	0.00	0.00	0.00	1.00[a]	0.00	0.99[a]	0.00
ATSI	0.83[b]	0.07	0.83[b]	0.07	0.86[c]	0.07	0.89	0.82	0.79[c]	0.10	0.80	0.21

[a]significant at the 1% level; [b]significant at the 5% level; [c]significant at the 10% level; [d]adapted from Villano et al. (2018). Significant values shown as 1 are due to rounding. The actual value is not equal to 1.

The second demographic (*D*) variable that affects student retention is age, which is consistently significant throughout our study. Furthermore, except for the OLS model, age and student retention show a non-linear relationship and vary over time. According to the OLS estimate, each additional year of age raises the length of enrolment by 0.22 weeks. To put these results into perspective, the difference in the expected length of enrolment for a 28-year-old student (the average age of the case institution) and an 18-year-old student is 2.2 weeks of extra enrolment. Using a maximum-likelihood approach, Table 3 shows that the probability of completing initially rises but falls after the age of 43. According to the survival models, the hazard ratio of a student decreases up to 51 years of age, after which it starts to increase again. These results converge on a similar conclusion; namely, age is a significant determinant of student length of enrolment and retention rate. This finding is particularly relevant to institutions with diverse student cohorts in age distribution.

ATSI represents the third determinant of student retention. Table 3 shows that while ATSI is independent of the length of enrolment under the OLS estimation, it reduces the likelihood of discontinuing. The survival models suggest that ATSI students experience lower hazard ratios of between 14% and 17.5% relative to non-ATSI students. This decrease in the hazard ratio may stem from several sources, including the filter effect associated with the initial barriers to entry to university or additional support provided to ATSI students from a specialist centre on campus. Taken together, our results indicate the need to introduce different retention strategies for ATSI and non-ATSI students for the case study institution. In these regards, it is worthwhile to investigate further the nuances of student retention within the ATSI cohort.

3.1.2. Institutional Variables
Following Harrison et al. (2014, 2015) and Villano et al. (2018), our mixed-method approach captures the institutional (*I*) variables such as fees, prior studies, mode of enrolment, qualification type, and school. Table 4 presents the estimated coefficients associated with these institutional (*I*) variables.

The first institutional (*I*) variable is fee categories, which consist of domestic Higher Education Loan Program (HELP), domestic upfront fee-paying, and international fee students. This categorical variable uses the base case of a student accessing HELP to pay fees. The OLS model indicates that domestic fee-paying students are enrolled for 9.7 weeks more than domestic HELP students. Furthermore, these students have a higher chance of discontinuing. However, the survival models detect no noticeable differences in the retention rate between these two cohorts. As such, our results remain inconclusive, perhaps due to the small sample of domestic fee-paying students in the case institution. In contrast, international fee students have a significant impact on student outcome and retention in all models. Specifically, the OLS model indicates that international students enrol for 3.7 weeks longer than domestic HELP students. The maximum-likelihood approach shows that, compared to domestic HELP students, international students have a lower probability of discontinuing and a higher probability of completing their studies. The survival models show that international students initially report a lower hazard than domestic students, but this difference dissipates by the end of their second year of study.

Table 4. Estimates of the institutional variables

Model	Survival: short run[d] HR	SE	Survival: enduring HR	SE	Survival: long run HR	SE	OLS: weeks enrolled b	SE	Likelihood of discontinuing RRR	SE	Likelihood of completing RRR	SE
Domestic fee	0.56	0.32	0.56	0.32	0.51	0.29	9.67[b]	4.57	2.14[b]	0.81	0.50	0.41
	1.01[c]	0.01	1.01[c]	0.01	1.01[c]	0.01						
International fee	0.20[a]	0.06	0.18[a]	0.06	0.21[a]	0.06	3.70[a]	0.69	0.42[a]	0.07	4.16[a]	0.72
International fee $\times t^2$	1.00[a]	0.00	1.00[a]	0.00	1.00[a]	0.00						
International fee $\times t^3$	1.00[b]	0.00	1.00[a]	0.00	1.00[b]	0.00						
Prior studies	0.60[a]	0.10	0.59[a]	0.10	0.64[a]	0.11	6.40[a]	0.43	3.01[a]	0.18	6.53[a]	0.59
Prior studies $\times \ln(t)$	1.13[a]	0.05	1.13[a]	0.05	1.11[b]	0.05						
On campus	1.24[a]	0.06	1.22[a]	0.06	1.26[a]	0.06	−0.82[b]	0.41	1.50[a]	0.10	0.53[a]	0.07
On campus $\times t$	0.92[a]	0.01	0.96[a]	0.01	0.94[a]	0.01						
Diploma	0.99	0.11	1.01	0.11	0.96	0.11	2.95[c]	1.58	1.25	0.21	7.81[a]	2.25
Advanced diploma	0.74[b]	0.09	0.75[b]	0.09	0.67[a]	0.08	9.01[a]	0.99	1.49[a]	0.15	4.89[a]	0.98
Advanced diploma $\times t$	1.00[c]	0.00	1.00[c]	0.00	1.00[a]	0.00						
Bachelors (graduate entry)	0.72[a]	0.05	0.72[a]	0.05	0.72[a]	0.05	1.90[b]	0.75	0.67[a]	0.06	7.44[a]	1.40
Bachelors (honours)	0.57[a]	0.07	0.56[a]	0.07	0.54[a]	0.07	5.98[a]	1.07	0.39[a]	0.06	33.49[a]	6.34
School 1	1.16[b]	0.08	1.18[b]	0.08	1.13[c]	0.08	−1.67[b]	0.73	1.20[c]	0.11	0.15[a]	0.03
School 2	1.28[a]	0.09	1.30[a]	0.09	1.28[a]	0.09	−1.62[b]	0.67	0.99	0.09	0.44[a]	0.07
School 2 $\times 1/t$	0.01[a]	0.01	0.01[a]	0.01	0.01[a]	0.01						
School 3	1.20[a]	0.08	1.21[a]	0.08	1.19[b]	0.08	−1.11[c]	0.67	1.30[a]	0.12	0.38[a]	0.07
School 4	1.31[a]	0.08	1.31[a]	0.08	1.31[a]	0.08	−3.14[a]	0.64	1.25[b]	0.11	0.34[a]	0.06
School 5	1.31[a]	0.09	1.32[a]	0.09	1.33[a]	0.09	−3.56[a]	0.75	1.22[b]	0.12	0.03[a]	0.01
School 6	0.89	0.07	0.90	0.07	0.87[c]	0.07	−2.66[a]	0.64	0.97	0.08	0.34[a]	0.06
School 6 $\times t$	1.00[a]	0.00	1.00[a]	0.00	1.00[a]	0.00						
School 7	0.69	0.26	0.67	0.25	0.82	0.31	26.01[a]	2.28	0.91	0.51	1.07	0.88
School 8	1.11	0.09	1.12	0.09	1.09	0.09	−0.95	0.75	1.41[a]	0.16	0.85	0.14
School 9	0.95	0.08	0.96	0.08	0.93	0.08	1.29	0.89	0.91	0.10	0.52[a]	0.11

[a] significant at the 1% level; [b] significant at the 5% level; [c] significant at the 10% level; [d] adapted from Villano et al. (2018). Significant values shown as 1 are due to rounding. The actual value is not equal to 1.

Our study classifies prior studies as the third institutional (*I*) variable. However, since this variable captures the post–high school education opportunities the student has already undertaken, it should be classified as part of the demographic (*D*) variable. Regardless of classifications, we expect a lower hazard ratio for students who had undertaken the previous study after the school-leaving age as they are better prepared for university. Our expectation is supported by the OLS results, where students with prior studies are enrolled for an additional 6.4 weeks. However, the maximum-likelihood approach shows that students with prior studies are more likely to both discontinue and complete their studies. We offer two explanations for this apparent contradiction. First, the rising opportunity cost associated with study for students with prior studies may deter them from pursuing further study if they already have a qualification. And second, students with prior studies have had experience with studying that better prepares them to complete their chosen courses. Against this backdrop, including prior studies as an explanatory variable needs to be treated with caution. Specifically, if used to explain retention, we must have information that allows the separation of these two countervailing effects. The survival models indicate that, initially, there is a lower hazard of discontinuing. This effect changes over time, with the gap between students with and without prior studies disappearing by week 35. Our result shows that the benefit of having previously studied only remains relevant within the first one to two teaching periods.

With respect to the mode of enrolment, Table 4 compares on- and off-campus students. Our results in all models indicate that this variable significantly affects student outcomes and retention. The OLS estimate indicates that on-campus students are enrolled for 0.8 weeks less than their off-campus counterparts. The maximum-likelihood approach shows that on-campus students are more likely to discontinue, and less likely to complete, their studies. The survival models report that the hazard ratio is only higher for on-campus students than for off-campus students for the first 12 weeks or one teaching period. Figure 3 shows that the hazard ratio continues to decline over time, indicating that off-campus students are at a higher risk of discontinuing beyond the first teaching period. These results highlight the need to adjust retention strategies between the student population by mode of enrolment over time. For the case study institution with a strong history of supporting off-campus students, our results from the survival models identify a gap in retention between on-campus and off-campus students.
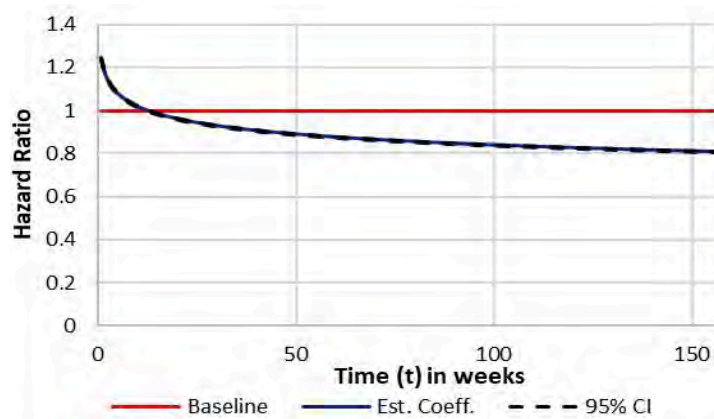

Figure 3. Estimates of on-campus student hazard over time.

The next variable of interest is the type of qualification undertaken by students, differentiating between diplomas, advanced diplomas, bachelor's degrees, bachelor's degrees with admission via graduate entry, and bachelor's degrees with honours. The base case for comparison is students who have graduate entry into a bachelor's degree. Comparing diploma students to bachelor's students, the OLS model indicates that the former stay enrolled for 2.95 weeks more than the latter. Meanwhile, advanced diploma students remain enrolled for a further 9 weeks. Advanced diploma students are more likely to both discontinue and complete their studies than bachelor's students, reflecting differences in the level of commitment in attaining qualifications. The maximum-likelihood approach shows that diploma students are more likely to complete their studies than bachelor's students. The survival models show no significant difference in the hazard of discontinuing. We attribute this result to fewer units required to complete a diploma than a bachelor's course. However, these students exhibit a lower hazard ratio over time, with a linear temporal effect indicating that the hazard of discontinuing rises over time for advanced diploma students. Similarly, students in graduate entry and honours awards are enrolled for an average of 1.9 and 6 weeks longer than bachelor's students. Both are less likely to discontinue, and more likely to complete, their qualifications. The survival model

also shows that these students report a lower hazard of discontinuing throughout their studies. Overall, our results suggest that students admitted through graduate admission or undergoing honours degrees are at a lower risk of discontinuing.

The final institutional (I) variable is schools in the case study institution. Since schools is a categorical variable, our interpretations are made in comparison to a base-case school. We expect variations between schools because of the different objectives and courses taught across various disciplines. Comparing all schools by model, the OLS results show that School 7 students are enrolled for an average of 26 weeks longer than the base case. Meanwhile, School 1 to School 6 show that students are enrolled for an average of between one and three weeks longer than the base case, but School 8 and School 9 report no difference in the enrolment period. For the maximum-likelihood models, the approach controls for significant differences between schools, which are then factored into the likelihood estimates. With respect to the survival models, School 1 to School 6 have a higher hazard of discontinuing. School 2 and School 6 stand out because of their temporal effects. In the case of School 2, the combined effect of the estimated coefficients for both the school (School 2) and the time (School 2 × 1/t) results in an overall insignificant difference in the hazard of discontinuing over time. Students in School 6 initially have a decreased hazard of discontinuing, but by week 29, their hazard ratio has increased linearly to the same level as the base-case school. The effect continues to strengthen over time, where students in School 6 face an increasingly higher hazard of discontinuing. This trend indicates scopes and reasons to develop targeted support strategies within School 6 to address student retention past the first year of study.

3.1.3. Student Performance and Study Load Variables

Student performance (P) and study load (L) represent significant variables when analyzing student retention. Table 5 shows the results for both student performance and study load, with most estimates exhibiting a temporal effect.

The first thing to note is that, independent of each grade obtained, the OLS model shows that the student, on average, enrols for a longer period. However, this result does not contribute to the overall interpretation in a mixed-method approach. Instead, the maximum-likelihood estimates in Table 5 indicate that only the grade outcome of fail incomplete increases the likelihood of discontinuing. The grade is given to students who failed to submit or complete assessment requirements within a unit, and as such it provides a strong indicator of risk of discontinuing. Similar to the OLS model, the other student performance variable captures student progression as much as it captures the result of any particular grade. Put differently, any grade other than a fail incomplete supports student progression and reduces the likelihood of discontinuing. A grade of withdrawn, withdrawn early, fail incomplete, or fail indicates that a student is less likely to complete their studies. Conversely, pass, credit, distinction, and high distinction grades improve students' likelihood of completing their studies. These results are consistent with our a priori expectations, where positive grade attainment that contributes to a student's attaining a degree will improve the chances of completion. The survival model provides a clearer picture of how grades affect the risk of discontinuing over time. Specifically, students who record a withdrawn or withdrawn early grade are at the most significant risk of discontinuing their first year of study. However, the temporal effect indicates that this risk decreases in magnitude to a point where a relatively small increase in the hazard of discontinuing occurs in the second year of study. This captures the point where students benefit from "gaming" the system; for instance, students will enrol in a unit of study before deciding whether the unit is worth pursuing or using their time and effort in alternative units.

The most complex of the grade results is fail incomplete. To account for TVCs and the proportional assumptions test used in survival models, we capture the temporal effect as a fourth- or fifth-order polynomial. Figure 4 presents the short-run results with a bimodal distribution function over time. Our result shows that the impact of fail incomplete on a given student's risk of discontinuing depends on when they receive the grade. Initially, the hazard of discontinuing increases, peaking at week 18, after which it decreases to a low point at around week 96, before peaking again in week 149. Importantly, throughout the sample period, it increases from obtaining fail incomplete, albeit only the magnitude of the effect varies over time.

Table 5. Estimates of student performance and load variables

Model	Survival: short run[d]		Survival: enduring		Survival: long run		OLS: weeks enrolled		Likelihood of discontinuing		Likelihood of completing	
	HR	SE	HR	SE	HR	SE	b	SE	RRR	SE	RRR	SE
Withdrawn	2.13[a]	0.45	1.80[a]	0.38	2.23[a]	0.48	0.09[a]	0.00	0.74[a]	0.02	0.49[a]	0.05
Withdrawn × t	1.01[b]	0.00	1.01[c]	0.00	1.01[b]	0.00						
Withdrawn × $t^{0.5}$	0.85[a]	0.05	0.88[b]	0.05	0.85[a]	0.05						
Withdrawn early	1.47[a]	0.08	1.45[a]	0.08	1.50[a]	0.09	0.13[a]	0.00	0.87[a]	0.01	0.59[a]	0.03
Withdrawn early × t	0.99[a]	0.00	0.99[a]	0.00	0.99[a]	0.00						
Withdrawn early × t^2	1.00[a]	0.00	1.00[a]	0.00	1.00[a]	0.00						
Fail incomplete	1.51[a]	0.06	1.39[a]	0.06	1.68[a]	0.12	0.10[a]	0.00	1.12[a]	0.02	0.57[a]	0.07
Fail incomplete × t	1.02[a]	0.00	1.02[a]	0.00	1.02[a]	0.00						
Fail incomplete × t^2	1.00[a]	0.00	1.00[a]	0.00	1.00[a]	0.00						
Fail incomplete × t^3	1.00[a]	0.00	1.00[a]	0.00	1.00[a]	0.00						
Fail incomplete × t^4	1.00[a]	0.00	1.00[a]	0.00	1.00[c]	0.00						
Fail incomplete × t^5					1.00	0.00						
Fail	1.14[a]	0.02	1.12[a]	0.02	1.15[a]	0.02	0.09[a]	0.00	0.94[b]	0.02	0.55[a]	0.04
Pass	0.85[a]	0.03	0.82[a]	0.03	0.88[a]	0.03	0.08[a]	0.00	0.71[a]	0.01	1.14[a]	0.02
Pass × t[e]	1.00	0.00	1.00	0.00	1.00	0.00						
Credit	0.87[a]	0.03	0.84[a]	0.03	0.89[a]	0.03	0.08[a]	0.00	0.72[a]	0.01	1.12[a]	0.02
Credit × t	1.00	0.00	1.00	0.00	1.00	0.00						
Distinction	0.86[b]	0.06	0.82[a]	0.06	0.90	0.07	0.09[a]	0.00	0.71[a]	0.01	1.09[a]	0.02
Distinction × t[e]	1.00	0.00	1.00	0.00	1.00	0.00						
Distinction × t^2[e]	1.00	0.00	1.00	0.00	1.00	0.00						
High distinction	0.83[c]	0.08	0.80[b]	0.08	0.86	0.09	1.00[a]	0.00	0.73[a]	0.02	1.13[a]	0.02
High distinction × t[e]	1.00	0.00	1.00	0.00	1.00	0.00						
High distinction × t^2[e]	1.00	0.00	1.00	0.00	1.00	0.00						
Other	0.74[a]	0.05	0.74[a]	0.05	0.74[a]	0.05	0.07[a]	0.00	0.55[a]	0.05	0.72[a]	0.05
Units enrolled	0.81[a]	0.03	0.81[a]	0.03	2.62[a]	0.19	−31.09[a]	0.40	1.33[a]	0.04	2.74[a]	0.17
Units enrolled × t	0.98[a]	0.00	0.98[a]	0.00								
Units enrolled × t^2	1.00[a]	0.00	1.00[a]	0.00								
Units enrolled × ln(t)					0.60[a]	0.01						

[a]significant at the 1% level; [b]significant at the 5% level; [c]significant at the 10% level; [d]adapted from Villano et al., (2018); [e]only for time (t) > 12 weeks. Significant values shown as 1 are due to rounding. The actual value is not equal to 1.
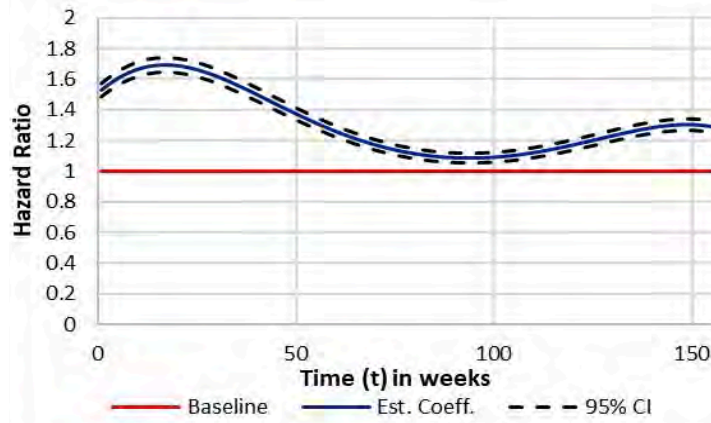
Figure 4. Estimates of fail incomplete over time.

The fail grade occurs when a student fails to attain a grade sufficient to pass the unit of study. Unlike fail incomplete, the hazard of discontinuing when achieving a fail grade is constant over time. Each time a student attains one of these grade outcomes, the hazard of discontinuing increases by between 12.1% and 14.9%.

Finally, the grades of pass, credit, distinction, and high distinction contribute to the students' progression, thus lowering the hazard of discontinuing. However, this effect is greater for pass and credit than distinction and high distinction. This is in line with the expectations around the relationship between performance and retention, where better-performing students are not at risk of discontinuing. In contrast, struggling students may deem achieving pass or credit a significant achievement, encouraging them to persevere.

The final set of variables of interest is the study load (L) variables. The OLS and maximum-likelihood models are captured by a weighted average of the number of units a student is enrolled in during any given week throughout their enrolment. Using the OLS model, Table 5 shows that a heavier study load reduces the length of enrolment. In fact, each additional unit of study reduces enrolment by approximately 31 weeks. This result is consistent because students who undertake more units tend to complete their qualifications sooner. The maximum-likelihood models suggest that raising the study load increases the likelihood of both discontinuing and completing the course. This result reflects the trade-off between the stress from enrolling in more units and reducing opportunity costs by completing their courses sooner.

In the survival models, the study load (L) variables differentiate between the short-run, enduring, and long-run models. Figure 5(a) presents the hazard of discontinuing the short-run model, with the enduring models indicating an almost identical pattern. The short-run and enduring models show that for each additional unit of study undertaken by a student, the hazard of discontinuing decreases. The effect is more pronounced as time goes by, too, indicating that the more students undertake in their later weeks of study, the more they will decrease the hazard of discontinuing.
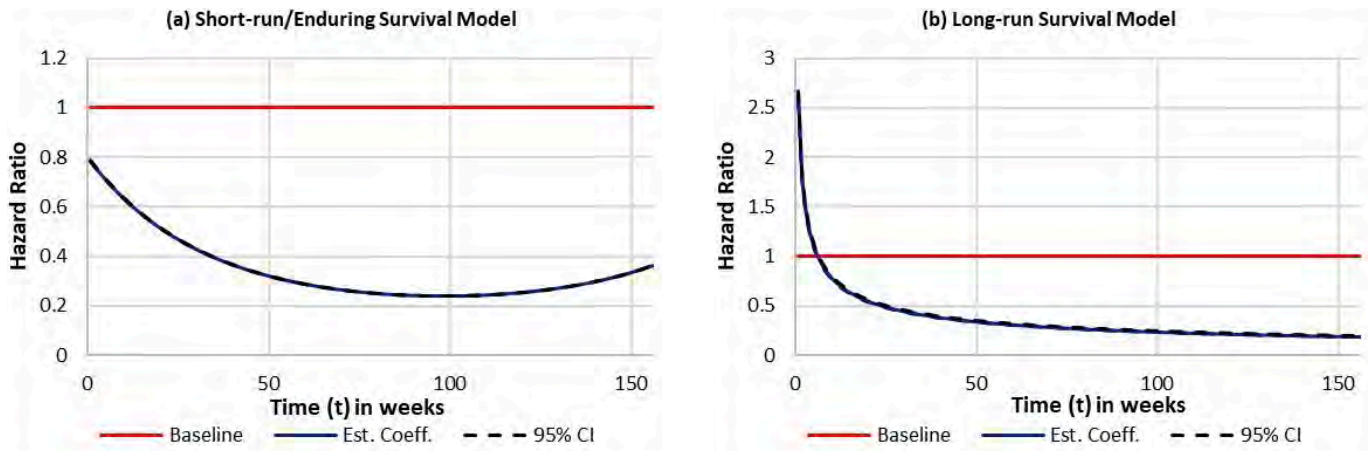

Figure 5. Estimates of study load over time.

Meanwhile, the long-run model in Figure 5(b) shows an initial increase in the hazard of discontinuing for the first seven weeks. The hazard of discontinuing increases with the number of units enrolled in. However, this effect decreases rapidly over time, converging to the short-run and enduring models. This initial spike in the long-run model may indicate the challenges facing students upon the initial enrolment and commencement of courses, with heavier study load compounding stress levels.

3.2. EAS and Student Retention

Our study's key focus resides in the effect of EASs on student retention and how much effect can be identified using a mixed-method approach. An important caveat here is that the EAS data only captures the identification of at-risk students, with no information on the post-identification engagement with student-support services. Consequently, the estimations cannot separate the effect of the EAS identifying a student who then chooses to access support services from the effect of the EAS identifying a student who chooses not to access support services. While this inference provides the strongest evidence on the EAS–student retention nexus, it does not preclude us from estimating the overall effect.

Using three years of observations, we use cross-sectional and panel data analyses to examine the effect of EAS on retention. We discuss these results to highlight the issues with identifying students in later years of enrolment and the limitations of temporal data modelling.

3.2.1. Overall Model Significance

All the models estimated used 16,091 students enrolled over a three-year period to form the base dataset. Table 6 summarizes the overall results for the estimated models. Using the p-value of the estimated coefficients as the test statistic, all models are found to be statistically significant at the 1% level. For the OLS model, the R^2 value indicates that 80.7% of the variation in the students' length of enrolment can be explained by the variation in the explanatory variables.

Table 6. Overall model results

Model	OLS	Likelihood	Survival: short run	Survival: enduring	Survival: long run
n	16091	16091	16091	16091	16091
DF	32	64	56	57	59
R^2/pseudo R^2	0.807	0.2747			
Root MSE	18.837				
F	1693.36				
p	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
LR chi^2		7431.43	7329.9	7324.87	7679.19
Log likelihood			−42279.97	−42282.48	−42105.33

3.2.2. Linking EAS Identification and Severity to Retention

Although cross-sectional data does not include temporal dimensions of the identification–discontinuation nexus, we can still extract useful information. In addition to the OLS model, we estimate the length of enrolment by the STE model, which measures the risk level by categorizing students according to the number of times the EAS identified them. Table 7 indicates the breakdown of 16,091 students in the dataset into five different risk categories, ranging from never identified by EAS to being identified more than 20 times throughout their enrolment. Table 8 provides the estimated coefficients with respect to both an overall EAS effect and a breakdown based on the EAS risk levels.

Table 7. Number of times identified

EAS risk level	Number of times identified	Frequency	Percent (%)
0	0	4,830	29.96
1	1–4	5,582	34.62
2	5–9	2,829	17.55
3	10–19	1,858	11.52
4	20+	1,025	6.36

Table 8. Summary of effects for the EAS

Model	STE: weeks enrolled		OLS: weeks enrolled		Likelihood of discontinuing		Likelihood of completing	
	ATE	*SE*	*b*	*SE*	*RRR*	*SE*	*RRR*	*SE*
POM / Constant	60.72[a]	1.24	63.00[a]	1.80	1.25[c]	0.15	0.01[a]	0.00
EAS	26.02[a]	2.34	8.96[a]	0.44	1.06	0.05	0.71[a]	0.08
EAS risk level 1	3.84[b]	1.82	8.93[a]	0.44	1.05	0.05	0.75[b]	0.09
EAS risk level 2	14.40[a]	3.41	13.12[a]	0.51	1.16[c]	0.09	0.44[a]	0.07
EAS risk level 3	79.13[a]	2.70	9.38[a]	0.67	1.54[a]	0.17	0.31[a]	0.06
EAS risk level 4	73.75[a]	2.64	4.99[a]	0.81	1.14	0.16	0.20[a]	0.05

[a]*significant at the 1% level;* [b]*significant at the 5% level;* [c]*significant at the 10% level*

Our OLS and STE results indicate that students identified by the EAS increase their enrolment by an average of 26 and 9 weeks, respectively. Considering the risk level, the STE model indicates that for at-risk level 1, students are enrolled for an additional 3.8 weeks; for at-risk level 2, it increases to an additional 14.4 weeks of enrolment. At the higher risk level 3, enrolment reaches its maximum of 79 weeks, and at highest risk level 4, it is 73.7 weeks of enrolment. This varies significantly from the OLS results, which indicated that for at-risk levels 3 and 4, students are only enrolled for an additional 9.4 and 5 weeks, respectively. This result captures a key difference between these two approaches. In part, these differences in the estimates can be attributed to the fact that the OLS estimates capture the effects based on the pooled samples for each group. In contrast, the STE estimates encapsulate the effects through a counterfactual approach that compares those students identified by the EAS to similar students who have never been identified by the EAS. In practice, the latter provides more reliable estimates on the actual effect of the EAS. However, interpreting these results becomes a little more difficult. The increased length of enrolment could be attributed to the student being retained longer at the university. However, it could also be attributed to the model estimating that students who are at higher hazard of discontinuing will need significantly more time to complete their studies. Indeed, this highlights a key limitation of the data available, where it is impossible to disaggregate these two plausible explanations.

Comparing the likelihood approach results, there is only a statistically significant increase in the likelihood of discontinuing at the risk level 3 category. All other categories have a weak or no relationship to discontinuing. As such, the likelihood approach does not seem to yield as much meaningful information about the effect of the EAS on student retention as the other models. However, the results for the likelihood of completing do make sense in this context. Specifically, the likelihood of completing a degree is significantly less than one for all levels. Importantly, the likelihood of completing decreases as the risk category rises. This negative relationship is important in highlighting that the likelihood approach may contribute to our understanding of the completion side of the retention equation more than the discontinuing likelihood model.

3.2.3. The EAS and Student Retention over Time

One advantage of using the survival models is capturing the temporal relationships between EAS identification and the discontinuation of the study. Table 9 shows that all models report a significant relationship across all periods, with Figure 6 showing the combined hazard based on the temporal estimates. Specifically, the enduring and long-run models can be compared to Villano et al. (2018) in Figure 6(a) for identifying how varying model specifications allow for different interpretations.

Figure 6(b) shows that the enduring model predicts a quadratic relationship over the time of a student's enrolment. To put it in context, although this effect starts like the short-run model, it weakens over time, with a minimum around week 90, before peaking again. In other words, it indicates that a student first identified by the EAS in the opening weeks of the teaching period will continue to have a higher hazard than a student identified around week 90 of their studies. However, the enduring model also highlights the hazard of being identified for the first time at a later stage of a student's enrolment. Being identified for the first time in week 120 indicates that the student has a significantly higher hazard than a student first identified in week 90. Therefore, getting a correct early identification plays a pivotal role when designing and deploying EASs and other support strategies within the university.

Table 9. Estimates of EAS: Panel data

Model	Survival: short run[d]		Survival: enduring		Survival: long run	
	HR	SE	HR	SE	HR	SE
EAS	2.21[a]	0.23	1.89[a]	0.20	0.02[a]	0.01
EAS × t	0.99[b]	0.00	0.99[a]	0.00	1.52[a]	0.06
EAS × t^2			1.00[b]	0.00	0.98[a]	0.00
EAS × t^3					1.00[a]	0.00
EAS × t^4					1.00[a]	0.00
EAS × t^5					1.00[a]	0.00
EAS × t^6					1.00[a]	0.00

[a]*significant at the 1% level;* [b]*significant at the 5% level;* [c]*significant at the 10% level;* [d]*adapted from Villano et al., (2018). Significant values shown as 1 are due to rounding; the actual value is not equal to 1.*
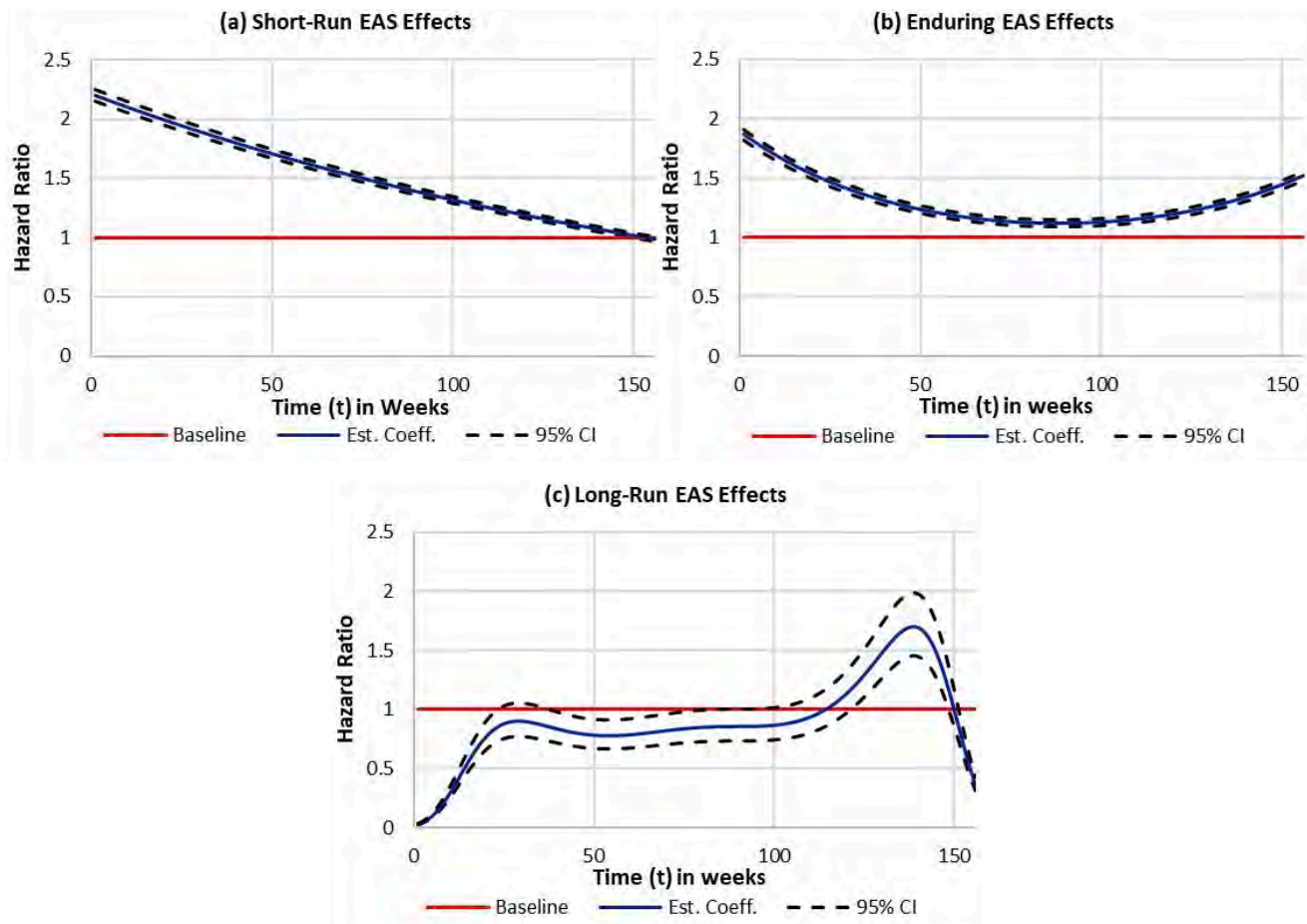

Figure 6. Estimates of EAS effects over time.

The long-run model provides an even more interesting result. Specifically, it divides students into those not identified by the EAS during their studies and those identified at some point during their studies, with the former group being used as the base case. To satisfy the proportional hazards assumptions test, a sixth-order polynomial was used to capture the TVCs.

The multi-modal graph in Figure 6(c) indicates that the not-identified group had a higher hazard of discontinuing their studies in the first 23 weeks of studies than the identified group. However, there are minor differences between these two groups until week 115, where the identified group has a significantly higher hazard of discontinuing than the not-identified group. This effect peaks at week 13, before decreasing rapidly again. With respect to the initial difference up to week 23, the students in the not-identified group are experiencing an increased number of discontinuation events during this period than the identified group. As such, the increased hazard for the not-identified group is possibly capturing the EAS's inability to identify students sufficiently early in their enrolment to place them in the identified group for support.

The later hump peaking around week 139 captures an underlying feature of the dataset. The underlying dataset captured several different enrolment outcomes, but for statistical modelling, the lapsed and discontinued student outcomes were pooled together (see Figure 1). Lapsed students occur when they have failed to enrol in any units of study for two years and have not formally contacted the university to discontinue their studies. As such, the hump occurring at week 139 may reflect the lapsed students exiting the system after this two-year period of inactivity. For the case study university, this result indicates the need to focus on the initial six months of student enrolment to improve retention.

4. Conclusion

To address the first research question, demographic, institutional, student performance, and study load variables were analyzed. All three demographic variables, gender, age, and ATSI status, show significant student retention relationships. The institutional variables, international fee-paying, on-campus, course type, and school, establish a consistent considerable relationship to student retention. Other institutional variables of domestic fee-paying and prior studies show varying levels of significance through the study. Although these variables may not affect retention, it can be argued that they need to be included in the models as control variables. Student performance measured by the grade distribution formed an important component of the models. These variables are significant in all models. The study load is a significant factor contributing to students and their decision to discontinue. However, our finding represents a double-edged sword for student support services when advising students whether they should decrease or increase their workload. On the one hand, increasing the workload may help students complete more quickly, reduce their risk of discontinuing, and become more actively engaged in their studies. On the other hand, doing so may result in less time for other activities, possibly compounding stresses associated with the study. This highlights a key area for future research.

The second research question concerns the effect of the EAS on student retention. In general, our results show that the EAS exerts a significant effect on retention, with the results from previous and new approaches indicating the following: (1) the EAS can identify students at risk of discontinuing; (2) the EAS is linked to a higher likelihood of completion if identified; (3) using the EAS as a treatment variable, the student's total length of enrolment increased; (4) the EAS in the long run identifies key areas for improvement with regard to retention and EAS design; and (5) the EAS performs the best when the data is collected at the early stage of student enrolment. These findings provide evidence linking the efficacy of the EAS to improving student retention. The converging evidence provided by the mixed-method approach helps to consolidate a body of evidence that can be used to support the implementation of such programs at other institutions.

The final research question focuses on the most appropriate methods for analysis. For this, the survival models provide important information about temporal effects, which is not possible using cross-sectional analyses such as the likelihood of OLS models outlined. Meanwhile, the STE model provides a statistically more rigorous result on cross-sectional data. It has strengths in analyzing students where it would not be possible for ethical or practical reasons to have a control group. This is a model that needs to be explored further in retention studies. Selecting from the approaches outlined, the best approach for analysis of EAS systems given the data provided is the survival models. However, future studies should combine survival models, capturing temporal changes, while linking this to treatment effects for causal inferences. As such, an STE model will provide the best option for detailed retention analysis in future studies.

One limitation of the study is the inability to make a complete causal inference: the EAS is causing an improvement in retention. To do this, it is necessary to isolate the effect of the EAS further by knowing which students had opted in for support services, compared to those who were identified as at risk but did not access any support. While there can be significant ethical issues surrounding the privacy of student data in these regards, using de-identified data with a simple binary variable indicating "accessed support" would be sufficient to make reasonable causal inferences about the impact of the EAS on student retention while accounting for the effect of the ongoing student support process itself and maintaining privacy. Future avenues of study are also to be undertaken differentiating the factors that affect decisions to formally discontinue studies, versus students who fail to engage in any administrative processes and fade out in their enrolment. Another aspect of future research would be to include factors from learning design, such as how the course and unit pedagogical factors may also affect students' decisions

to continue their studies. To do so would also require data and information to be collected from the learning environment, which is outside the scope of this study.

The results presented have important implications for student retention analysis within the learning analytics community. First, the factors that affect student retention are complex, with many variables having temporal effects. The results demonstrate the need to have temporal approaches to analyzing retention, which is only possible if suitably granular data is available. This raises a second point, namely, the need for comprehensive data capturing the many aspects of the learning environment. The dataset provided by the case study institution allowed analysis to take place with unprecedented detail in the survival models. This supports other research indicating that "teachers, students, faculty, support staff, and administrators can all benefit by applying data to understand what's happening in classrooms and how to improve and optimize learning" (Siemens et al., 2013). Finally, the study helps benchmark and demonstrate what can be achieved by quantitative analysis methods, supporting the learning analytics community's central objective of an evidence-based understanding of the learning environment.

Declaration of Conflicting Interest

Funding

Acknowledgments

References

Aguilar, S., Lonn, S., & Teasley, S. D. (2014). Perceptions and use of an early warning system during a higher education transition program. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 113–117). New York: ACM. https://doi.org/10.1145/2567574.2567625

Arnold, K. E., Tanes, Z., & King, A. S. (2010). Administrative perceptions of data-mining software signals: Promoting student success and retention. *The Journal of Academic Administration in Higher Education, 6*(2), 29–39.

Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel, 25*(4), 297–308.

Bean, J. P. (1980). Dropouts and turnover: The synthesis of a causal model of student attrition. *Research in Higher Education, 12*(2), 155–187. https://doi.org/10.1007/BF00976194

Bean, J. P., & Metzner, B. S. (1985). A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research, 55*(4), 485–540. https://doi.org/10.3102/00346543055004485

Department of Education. (2014). Data from: Selected Higher Education Statistics — 2013 Student Data, All Students Table (Dataset). Department of Education. Canberra. Accessed 20 June 2015. http://docs.education.gov.au/system/files/doc/other/2013allstudents_0.xls

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. https://www.jstor.org/stable/1558433

Ferguson, R., & Clow, D. (2017). Where is the evidence? A call to action for learning analytics. In *Proceedings of The Seventh International Conference on Learning Analytics and Knowledge* (LAK '17), 13–17 March 2017, Vancouver, BC, Canada (pp. 56–65). New York: ACM. https://doi.org/10.1145/3027385.3027396

Harrison, S., Villano, R., Lynch, G., & Chen, G. (2014). Timing student departure by estimating the length of enrolment for discontinuing students. Paper presented at the 2014 Australian Learning Analytics Summer Institute, University of Technology Sydney, 20–21 November 2014.

Harrison, S., Villano, R., Lynch, G., & Chen, G. (2015). Likelihood analysis of student enrollment outcomes using learning environment variables: A case study approach. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 141–145). New York: ACM. https://doi.org/10.1145/2723576.2723621

Harrison, S. (2016). *Microeconometric analysis of the relationships between early alert systems and student retention*. PhD dissertation, University of New England, Armidale, NSW, Australia.

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* (LAK '16), 25–29 April 2016, Edinburgh, UK (241–248). New York: ACM. https://doi.org/10.1145/2883851.2883923

Ishitani, T. T. (2003). A longitudinal approach to assessing attrition behavior among first-generation students: Time-varying effects of pre-college characteristics. *Research in Higher Education, 44*(4), 433–449. https://doi.org/10.1023/A:1024284932709

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. https://doi.org/10.1080/00221546.2006.11778947

Ishitani, T. T., & DesJardins, S. L. (2003). A longitudinal investigation of dropout from college in the United States. *Journal of College Student Retention: Research, Theory & Practice, 4*(2): 173–201. https://doi.org/10.2190/V4EN-NW42-742Q-2NTL

Jayaprakash, S. M., Moody, E.W., Lauría, E. J. M., 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. https://doi.org/10.18608/jla.2014.11.3

Jones-White, D. R., Radcliffe, P. M., Huesman, R. L., & Kellogg, J. P. (2010). Redefining student success: Applying different multinomial regression techniques for the study of student graduation across institutions of higher education. *Research in Higher Education, 51*(2): 154–174. https://doi.org/10.1007/s11162-009-9149-4

Lin, T.-C., Yu, W. W.-C., & Chen, Y.-C. (2012). Determinants and probability prediction of college student retention: New evidence from the probit model. *International Journal of Education Economics and Development, 3*(3), 217–236. http://doi.org/10.1504/IJEED.2012.049174

Marrington, A. D., Nelson, K. J., & Clarke, J. A. (2010). An economic case for systematic student monitoring and intervention in the first year in higher education. In *Proceedings of the 13th Pacific Rim First Year in Higher Education Conference*, 27–30 June 2010, Adelaide, Australia (pp. 1–4). Brisbane, Australia: Queensland University of Technology Publications. Retrieved from https://eprints.qut.edu.au/33231/

Nelson, K. J., & Creagh, T. (2013). *A Good Practice Guide: Safeguarding Student Learning Engagement*. Brisbane: Queensland University of Technology. Retrieved from https://eprints.qut.edu.au/59189/

Parnell, A., Jones, D., Wesaw, A., & Brooks, D. C. (2018). *Institutions' use of data and analytics for student success: Results from a national landscape analysis*. EDUCAUSE Center for Analysis and Research (ECAR). Retrieved from https://www.naspa.org/images/uploads/main/DATA2018_DOWNLOAD.pdf

Siemens, G., Dawson, S., & Lynch, G. (2013). *Improving the quality and productivity of the higher education sector: Policy and strategy for systems-level deployment of learning analytics*. Sydney, Australia: Office for Learning and Teaching. Retrieved from https://solaresearch.org/wp-content/uploads/2017/06/SoLAR_Report_2014.pdf

Simons, J. M. (2011). *A national study of student early alert models at four-year institutions of higher education*. ProQuest LLC. Retrieved from https://eric.ed.gov/?id=ED535792

Singell, L. D., & Waddell, G. R. (2010). Modeling retention at a large public university: Can at-risk students be identified early enough to treat? *Research in Higher Education, 51*(6), 546–572. https://doi.org/10.1007/s11162-010-9170-7

Stratton, L. S., O'Toole, D. M., & Wetzel, J. N. (2007). Are the factors affecting dropout behavior related to initial enrollment intensity for college undergraduates? *Research in Higher Education, 48*(4), 453–485. https://doi.org/10.1007/s11162-006-9033-4

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research, 45*(1), 89–125. https://doi.org/10.3102/00346543045001089

Tinto, V. (2006). Research and practice of student retention: What next? *Journal of College Student Retention: Research, Theory & Practice, 8*(1), 1–19. https://doi.org/10.2190/4YNU-4TMB-22DJ-AN4W

Villano, R., Harrison, S., Lynch, G., & Chen, G. (2018). Linking early alert systems and student retention: A survival analysis approach. *Higher Education, 76*, 903–920. https://doi.org/10.1007/s10734-018-0249-y

Appendix: Survival Analysis Output

Table A1. Proportional hazards assumptions test (short-run model)

Variables	rho	chi^2	$p >$ chi^2	Variables	rho	chi^2	$p >$ chi^2
Gender ($f = 1$)	−0.01	0.43	0.51	Withdrawn	0.00	0.03	0.87
Age	0.00	0.03	0.86	Withdrawn × t	0.00	0.00	0.95
Age^2	−0.01	0.39	0.53	Withdrawn × $t^{0.5}$	0.00	0.01	0.92
ATSI	0.02	2.09	0.15	Withdrawn early	0.00	0.08	0.77
Domestic fee	0.00	0.00	0.97	Withdrawn early × t	−0.01	0.19	0.66
Domestic fee × t	0.00	0.00	0.98	Withdrawn early × t^2	0.01	0.26	0.61
International fee	−0.01	0.65	0.42	Fail incomplete	−0.02	1.18	0.28
International fee × t^2	0.01	1.08	0.30	Fail incomplete × t	−0.02	2.79	0.10
International fee × t^3	−0.01	1.05	0.31	Fail incomplete × t^2	0.02	2.93	0.09
Prior studies	0.02	2.31	0.13	Fail incomplete × t^3	−0.02	2.63	0.11
Prior studies × ln(t)	−0.03	2.88	0.09	Fail incomplete × t^4	0.02	2.45	0.12
On campus	0.01	0.49	0.49	Fail	−0.01	1.09	0.30
On campus × t	0.00	0.10	0.75	Pass	−0.02	2.47	0.12
Diploma	0.00	0.04	0.84	Pass × t*	0.02	2.46	0.12
Advanced diploma	0.02	1.10	0.29	Credit	−0.01	0.82	0.37
Advanced diploma × t	−0.01	0.56	0.46	Credit × t	0.01	0.56	0.45
Bachelors (graduate entry)	−0.01	0.13	0.72	Distinction	0.01	0.15	0.70
Bachelors (honours)	0.02	1.36	0.24	Distinction × t*	−0.01	0.27	0.60
School 1	0.01	0.51	0.48	Distinction × t^2*	0.01	0.29	0.59
School 2	−0.01	0.19	0.66	High distinction	0.00	0.00	0.99
School 2 × 1/t	0.01	0.21	0.65	High distinction × t*	0.00	0.02	0.89
School 3	0.00	0.04	0.83	High distinction × t^2*	0.00	0.07	0.79
School 4	−0.01	0.29	0.59	Other	0.00	0.08	0.78
School 5	0.01	0.41	0.52	Units enrolled	−0.01	0.57	0.45
School 6	−0.01	0.45	0.50	Units enrolled × t	0.01	0.61	0.43
School 6 × t	0.02	2.06	0.15	Units enrolled × t^2	−0.01	0.52	0.47
School 7	0.00	0.00	0.95	EAS	0.01	0.26	0.61
School 8	0.01	0.82	0.36	EAS × t	−0.01	0.43	0.51
School 9	0.00	0.06	0.80	Global test		30.74	1.00

* for $t > 12$

Table A2. Proportional hazards assumptions test (enduring model)

Variables	rho	chi^2	$p > \text{chi}^2$	Variables	rho	chi^2	$p > \text{chi}^2$
Gender ($f = 1$)	−0.01	0.44	0.51	Withdrawn	0.00	0.02	0.90
Age	0.00	0.03	0.87	Withdrawn $\times t$	0.00	0.01	0.91
Age^2	−0.01	0.35	0.56	Withdrawn $\times t^{0.5}$	0.00	0.01	0.91
ATSI	0.02	2.16	0.14	Withdrawn early	0.01	0.18	0.67
Domestic fee	0.00	0.00	1.00	Withdrawn early $\times t$	−0.01	0.23	0.63
Domestic fee $\times t$	0.00	0.00	0.96	Withdrawn early $\times t^2$	0.01	0.24	0.63
International fee	−0.01	0.60	0.44	Fail incomplete	−0.01	0.53	0.47
International fee $\times t^2$	0.01	0.83	0.36	Fail incomplete $\times t$	−0.02	2.81	0.09
International fee $\times t^3$	−0.01	0.78	0.38	Fail incomplete $\times t^2$	0.02	2.59	0.11
Prior studies	0.02	2.48	0.12	Fail incomplete $\times t^3$	−0.02	2.30	0.13
Prior studies $\times \ln(t)$	−0.03	3.06	0.08	Fail incomplete $\times t^4$	0.02	2.16	0.14
On campus	0.01	0.38	0.54	Fail	−0.01	0.15	0.70
On campus $\times t$	0.00	0.11	0.74	Pass	−0.01	0.61	0.43
Diploma	0.00	0.04	0.83	Pass $\times t$ *	0.01	0.46	0.50
Advanced diploma	0.02	1.07	0.30	Credit	0.00	0.07	0.79
Advanced diploma $\times t$	−0.01	0.56	0.46	Credit $\times t$	0.00	0.02	0.88
Bachelors (graduate entry)	−0.01	0.11	0.74	Distinction	0.01	0.30	0.58
Bachelors (honours)	0.02	1.24	0.27	Distinction $\times t$ *	−0.01	0.46	0.50
School 1	0.01	0.62	0.43	Distinction $\times t^2$ *	0.01	0.49	0.48
School 2	−0.01	0.23	0.63	High distinction	0.00	0.05	0.83
School 2 $\times 1/t$	0.01	0.28	0.59	High distinction $\times t$ *	0.00	0.08	0.78
School 3	0.00	0.06	0.80	High distinction $\times t^2$ *	0.01	0.12	0.73
School 4	−0.01	0.25	0.62	Other	0.00	0.02	0.88
School 5	0.01	0.38	0.54	Units enrolled	−0.01	0.76	0.38
School 6	−0.01	0.36	0.55	Units enrolled $\times t$	0.01	1.03	0.31
School 6 $\times t$	0.02	1.80	0.18	Units enrolled $\times t^2$	−0.01	1.01	0.31
School 7	0.00	0.00	1.00	EAS	0.00	0.05	0.82
School 8	0.01	0.72	0.40	EAS $\times t$	−0.01	0.11	0.74
School 9	0.00	0.03	0.86	EAS $\times t^2$	0.00	0.08	0.77
				Global test		26.96	1.00

* for $t > 12$

Table A3. Proportional hazards assumptions test (long run model)

Variables	rho	chi^2	$p >$ chi^2	Variables	rho	chi^2	$p >$ chi^2
Gender ($f = 1$)	−0.01	0.35	0.55	Withdrawn	0.00	0.02	0.88
Age	0.01	0.63	0.43	Withdrawn $\times t$	0.00	0.00	0.95
Age^2	−0.02	1.09	0.30	Withdrawn $\times t^{0.5}$	0.00	0.01	0.93
ATSI	0.02	1.28	0.26	Withdrawn Early	0.01	0.21	0.64
Domestic fee	0.00	0.00	0.97	Withdrawn Early $\times t$	−0.01	0.46	0.50
Domestic fee $\times t$	0.00	0.00	0.98	Withdrawn Early $\times t^2$	0.01	0.58	0.45
International fee	−0.01	0.87	0.35	Fail Incomplete	0.00	0.00	0.99
International fee $\times t^2$	0.02	1.61	0.21	Fail incomplete $\times t$	−0.01	0.79	0.38
International fee $\times t^3$	−0.02	1.58	0.21	Fail incomplete $\times t^2$	0.00	0.01	0.93
Prior studies	0.02	1.57	0.21	Fail incomplete $\times t^3$	0.00	0.02	0.89
Prior studies $\times \ln(t)$	−0.02	2.10	0.15	Fail incomplete $\times t^4$	0.00	0.07	0.79
On campus	0.01	0.19	0.67	Fail incomplete $\times t^5$	0.01	0.12	0.73
On campus $\times t$	0.01	0.12	0.73	Fail	−0.02	1.83	0.18
Diploma	0.00	0.00	0.99	Pass	−0.02	2.73	0.10
Advanced diploma	0.02	1.75	0.19	Pass $\times t$ *	0.03	3.37	0.07
Advanced diploma $\times t$	−0.02	1.51	0.22	Credit	−0.02	1.20	0.27
Bachelors (graduate entry)	0.00	0.07	0.79	Credit $\times t$	0.01	0.96	0.33
Bachelors (honours)	0.02	2.03	0.15	Distinction	0.00	0.06	0.81
School 1	0.01	0.77	0.38	Distinction $\times t$ *	−0.01	0.13	0.72
School 2	−0.01	0.17	0.68	Distinction $\times t^2$ *	0.01	0.13	0.72
School 2 $\times 1/t$	0.01	0.14	0.71	High distinction	0.00	0.02	0.88
School 3	0.00	0.09	0.77	High distinction $\times t$ *	0.00	0.00	1.00
School 4	−0.01	0.33	0.57	High distinction $\times t^2$ *	0.00	0.02	0.89
School 5	0.00	0.09	0.77	Other	0.00	0.05	0.82
School 6	−0.01	0.25	0.62	Units enrolled	−0.01	0.98	0.32
School 6 $\times t$	0.02	1.13	0.29	Units enrolled $\times \ln(t)$	0.01	1.17	0.28
School 7	0.00	0.01	0.92	EAS	0.00	0.00	0.99
School 8	0.02	1.49	0.22	EAS $\times t$	0.00	0.06	0.80
School 9	0.00	0.02	0.88	EAS $\times t^2$	−0.01	0.12	0.73
				EAS $\times t^3$	0.01	0.19	0.66
				EAS $\times t^4$	−0.01	0.25	0.61
				EAS $\times t^5$	0.01	0.31	0.58
				EAS $\times t^6$	−0.01	0.35	0.55
				Global test		30.81	1.00

* for t > 12