Helping students to persist with and succeed in their first year of higher education

Salah Kutieleh and Ewa Seidel
Transition Office, Flinders University

Attrition, particularly amongst students in their first year of study, has been problematic for universities because of economic and social ramifications. While differences exist in the way universities have approached this issue, the literature shows the need for a holistic strategy that aims to facilitate and enhance student engagement with their institutions. Flinders University has supported the development of a centrally run program to contact first-year students who are likely to drop out. Predictions from machine learning models, built using big data, were utilised to target intervention throughout 2014. Students were contacted by their peers, who welcomed and advised students of the services on campus. Almost two-thirds of the 4100 students in scope were successfully contacted. Preliminary findings showed an effective targeting strategy, with reduced student attrition, improved appreciation and effective utilisation of student support services. These findings continue to shape the ongoing development of the Student Success Program.

Background

Student attrition from higher education is a problem worldwide. It signifies a personal loss to the individual; it carries a social and economic cost to the community; and it results in a significant financial cost to each university. A Hobson Retention Project estimated that in 2008, the cost of attrition averaged out to a loss of $36 000 000 for each university (Adams et al., 2010). Attrition is particularly high among students in their first year as they transition to university and establish patterns of academic and social behaviour. Coates (2014) reports that the national attrition rate for commencing university students in Australia is around 13.5%, which is relatively low in comparison with other countries. However, some universities reported a higher attrition rate than this, including Flinders University, which is attributed partly to the fact that Flinders is over-represented by most equity groups that are known to have higher levels of attrition (Department of Education, 2013). In revenue terms, Adams et al (2010), estimate that dropping attrition rates by as little as 1% would equate to savings of up to $2 600 000 per university.

Research findings indicate that retaining students beyond the first semester increases a student’s long-term retention chances increases the likelihood of their academic success and, ultimately, increases their satisfaction rating of the institution (Kift, 2009). Tinto’s model (1993) and his subsequent work have frequently been used as the theoretical basis for many studies pertaining to student retention and attrition in higher education. Tinto argues that students’ background factors determine student expectations of higher education. Meeting such expectations will normally determine, to a large extent, if students persist with or depart from the university. As a direct consequence of this, engagement has become a ‘cornerstone of the tertiary education lexicon’ (Krause, 2007, p. 1). Student diversity in higher education has contributed to the complexity of student engagement. Researchers, administrators and university senior management consider it important to intervene before rather than after
students withdraw from higher education. Pearson & Naung (2013) maintain that intervention strategies aimed at students who are at risk of disengaging or failing would be most effective if implemented in a timely manner. However, successful intervention is dependent on the early identification of those students who are most at risk of withdrawing or failing.

Continual increases in variety and availability of data have allowed the higher education sector to consider the use of predictive analytics to identify students who are more likely to discontinue their university studies in their first year. The reported literature reveals that many universities, both in Australia and overseas, have recognised the need for a comprehensive data-driven approach as part of the arsenal against attrition (Siemen, Dawson & Lynch, 2013). Hanover (2014) recently reported a trend in the use of predictive analytics in American higher education institutions to support retention strategies. Nationally recognised programs at the Queensland University of Technology (QUT) and Griffith University, amongst others, are based on predictive analytics to identify at-risk students, implement intervention strategies and track students’ engagement with their studies and the institution during the various phases of their studies. Such programs have been shown to ‘significantly improve persistence and success’ (Nelson et al., 2009).

The accuracy and precision of predictive models tend to improve when more sources of data are considered; this further enhances the effectiveness of data-driven intervention strategies. Flinders University, the focus of this paper, has utilised a large and growing range of available data sources in machine learning approaches to predict each commencing student’s risk to ensure that students with high-risk probabilities are assisted to make an informed decision in relation to persisting with or discontinuing their studies.

After extensive consultation across the University and the sector, Flinders developed an analytics-based student support solution, labelled the Student Success Project (SSP), which was intended to identify students at risk through the use of emerging big data predictive analytics techniques. The aim was to provide support to such students, improve their chances of successful completion, and reduce attrition.

The pilot project, which focused on the 2014 first-year cohort, comprised:

- Commencing (first-year) students (including international students);
- Enrolled students in a Bachelor Pass degree or four year Bachelor Honours (direct entry);
- First time in higher education students;
- Actively enrolled students (i.e. not zero current study load).

Consistent with the transition timeline as proposed by Kift (2009), four campaigns were conducted in each semester, in weeks two, four, eight and twelve; these coincided with known critical stages in the students’ transition timeline. Each campaign consisted of targeting support towards the highest-risk students, offering timely advice and referring them onto support services relevant to the transition to university.

**Project Execution**

The project required collaboration between a number of key areas within the University, spanning both academic and professional ranks. It also required collaboration with other universities in an attempt to refine and improve technical and non-technical aspects of the project. Significant time and external and internal expertise were required to develop a
sophisticated intervention model that would achieve multifaceted contact points and additional support services. Importantly, a whole-of-University approach was implemented to simplify the communication process. This meant that students received a consistent message from the University. The value expressed by Tinto (2009) that ‘student retention is everyone’s business’ guided the approach used to contact students.

An exhaustive review of the relevant literature, and consultations with stakeholders at Flinders and at other universities led to the conceptualisation of the SSP, consisting of three broad areas as represented in Figure 1 below:

- Analytics—data-driven student insights (modelling);
- Infrastructure—a recording system to facilitate the desired business process (CRM);
- Process—student communication and follow-up strategy (intervention).

![Figure 1: Interaction between analytics and intervention with supporting infrastructure (CRM)](image)

The analytics area used SPSS Modeller to:

- Connect to databases across the University to extract, integrate and manipulate a variety of data sources;
- Develop a data-mining process;
- Build models to predict each student’s likelihood of attrition;
- Output student data and predicted risk values required for intervention.

As is demonstrated in Table 1 below, the predictive analytics component brought together a variety of data, including data for both current and previous first-year students.
<table>
<thead>
<tr>
<th>Data Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic Performance</td>
<td>ATARs, first-semester topic results and GPAs</td>
</tr>
<tr>
<td>Flinders Learning Online (FLO) Logins</td>
<td>Logins related to topic interactions online</td>
</tr>
<tr>
<td>South Australian Tertiary Admission Centre (SATAC)</td>
<td>Applications and offers into University</td>
</tr>
<tr>
<td>Student Enrolments</td>
<td>Demographic, topic and course information</td>
</tr>
<tr>
<td>Student System Logins</td>
<td>Logins related to enrolments into courses, topics, and classes</td>
</tr>
</tbody>
</table>

**Table 1: Data Sources Used in Predictive Analytics**

Enrolment and academic performance data were considered because, historically, Flinders students had differentiated attrition rates across elements of their enrolment and academic performance. SATAC application data was considered, as it contains several variables indicating an applicant’s willingness to pursue higher education. Student System and FLO logins are transactional type data that capture the behaviours of students in terms of engagement. Behavioural data such as these were chosen because they are commonly identified in the literature as strong predictors in business attrition models and in predictive models built for learning analytics processes (Hanover 2013).

A separate predictive model was developed for each campaign to allow for changes in student enrolment characteristics and availability of new data sources and variables. Each model captured the effect of different drivers of success and captured students who switched from being active and engaged learners to being at risk of withdrawing. Consequently, all students were assigned revised probability values at each campaign, to be used in revised communication strategies.

Associations between student-level characteristics and attrition were drawn from pre-existing historic student data and were used to develop Chi-Squared Automatic Iterative Detection (CHAID) decision tree models that were then applied to current commencing students to predict their current probability of early departure.

The CHAID model is a statistical tool that utilises chi-squared significance tests to automatically and iteratively divide the population up into subsets, according to student characteristics that best describe each student’s probability of early departure. CHAID works well when using large datasets; it can handle categorical data (the most common format of higher education data); and it deals well with missing information. The CHAID model was chosen because it is suitable, and simple to implement, understand and interpret.

Gains charts were used as a visual aid to assess the accuracy of each predictive model and to assess how much the SSP could gain by using the predicted probabilities to prioritise ranking for communication during each campaign. One of eight Gains charts created, for the semester two, week two campaign, is shown in Figure 2.
Ensuring similar shaped blue curves in both windows of each Gains chart was the main criterion to confirm that each predictive model was built on sound statistical grounds. Figure 2 shows that by contacting 20% of first-year students with the largest predicted probabilities, and thus following the blue curve, the SSP could reach 40% more students who would have dropped out, compared to only 20% of attrited students if chosen at random and thus following the red line. This additional 40% gain is evidence of an effective predictive model. This was the average across all SSP campaigns and was higher than most examples reported in the higher education sector.

**Communication Strategy**

Students’ predicted risk values were used to prioritise communication with students. Resource availability, extent of non-contact, accuracy of student contact details, amount of follow-up, and duration of conversations determined the number of students the SSP was able to contact.

Based on the practices reported by other universities in Australia and elsewhere, the project Steering Group recommended the use of what has been labelled the peer-to-peer contact model, where commencing students are contacted by their peers rather than by the University’s staff. Consequently, a team of ‘SSP Officers’, consisting of current students and recent graduates, was recruited on the basis of their communication skills and familiarity with the University’s processes and procedures. These officers undertook training to ensure they communicate a consistent message to students and demonstrate familiarity with the support services at all levels at Flinders. They were monitored and supervised by the Executive Officer of the project, who developed the communication strategy and supervised the day-to-day running of the program.

It was essential to ensure that student needs were addressed in a relevant and timely manner by referring students, when necessary, to support services. The Steering Group adjusted the emphasis of each campaign to reflect the foci of the students’ concerns. The focus of communication was tailored to the transition timeline of the academic year. The first two campaigns in semester one focused on ensuring a smooth transition to the University and minimising preventable early departures. The last two campaigns in semester one focused on academic success and exam preparedness. In addition to the

---

**Figure 2: Gains Chart for Semester Two, Week Two Predictive Model**

---
areas of focus in semester one, semester two campaigns encouraged tutorial attendance and participation in assessment tasks.

By offering timely advice and referring students onto support services relevant to the transition timeline and the student’s needs, it was hoped that the SSP would reduce a student’s real risk of early departure, and that the expected attrition rate would be lower than the model predicted.

**Findings**

The aim of the program was to support students with high-risk probabilities in making an informed decision in relation to their university studies. Just over 4100 first-year students were in scope of the SSP in 2014. This included students who were enrolled at the time of any of the campaigns.

Recognising the prematurity of measuring the annual attrition rate for students who were in scope of the SSP in 2014, an alternative early measure of attrition was considered. It loosely formed a subset of annual attrition and was the basis of the predictive models. This included a measure of the rate at which all students in scope of a particular campaign withdraw from all topics, and therefore their course, at the ensuing Census date.

Between the first campaign in week two of semester one and Census in semester two, 660 students in scope (16.1%) withdrew from all their topics and left Flinders University. The predictive model successfully identified these students, which made it possible for the SSP Officers to contact almost 80% of them. However, it is impossible to determine whether some of these students have applied to other universities or have left higher education altogether.

By the end of semester two, the SSP Officers contacted almost 70% of all commencing first-year students. The extensive coverage provided the Steering Group with additional intelligence that will be useful in the process of considering the range of risk probabilities to target in the future, as it may be impractical to target such a high proportion of students in the future.

To further validate the effectiveness of the predictive models, 100 students who had the lowest predicted risk probability in week eight of semester one were contacted and assessed for their risk of withdrawing. The SSP Officers were asked to rank on a five-point Likert scale their assessment of the likelihood of these students continuing with or withdrawing from the University. Their assessment confirmed that the majority of the predicted low-risk students had no intention of withdrawing, providing a qualitative validation of the model’s accuracy. Despite their lack of risk, the students were still happy to receive the phone call.

Data presented in Table 2 below presents semester-on-semester attrition and provides early evidence that attrition among first-year students is plateauing or even improving somewhat. This result is particularly positive given an increase in annual attrition since 2009. The slight decline in semester-on-semester attrition rates in 2014 equates to the University saving an additional 44 students when compared to the potential attrition rate represented by the 2013 semester-on-semester attrition rate.
Comparisons of attrition predicted by the CHAID models with actual semester-on-semester attrition measures suggests the SSP has had a positive effect on retention and that the tabulated preliminary attrition rates would have been higher without SSP intervention.

Preliminary attrition rates among students who were contacted also suggest that high-risk students who attended Orientation Week had a reduced rate of attrition. This information, combined with feedback from students and SSP Officers, suggests a need for a campaign before each semester begins to encourage students to attend Orientation Week.

Results from predictive model development showed the leading indicators of attrition differed at each campaign. Contrary to popular belief, the individual profile of a student was not a strong predictor of attrition. Instead, the leading indicators tended to be behavioural data related to learning and academic support. This finding meant that students were identified and contacted based on their patterns of behaviour rather than their profile, thus avoiding potentially stigmatising students through stereotyping.

Referrals were made to students when they either expressed a desire or need to address their concerns. Table 3 presents the top main referrals made during semester one. Additional analysis on semester two is referenced below but not tabulated.

<table>
<thead>
<tr>
<th>Services referred to</th>
<th>Semester One</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wk 2 Wk 4 Wk 8 Wk 12</td>
</tr>
<tr>
<td>Careers and Employer Liaison Centre</td>
<td>22% 20% 11% 5%</td>
</tr>
<tr>
<td>Student Learning Centre</td>
<td>21% 22% 65% 13%</td>
</tr>
<tr>
<td>Enrolment</td>
<td>3% 2% 3% 13%</td>
</tr>
<tr>
<td>Student Finances</td>
<td>0% 1% 0% 0%</td>
</tr>
<tr>
<td>FLO Help desk</td>
<td>3% 6% 1% 0%</td>
</tr>
<tr>
<td>Flinders University Student Association</td>
<td>13% 11% 10% 3%</td>
</tr>
<tr>
<td>Health and Counselling</td>
<td>13% 16% 11% 10%</td>
</tr>
<tr>
<td>Director of First Year Studies</td>
<td>8% 7% 3% 10%</td>
</tr>
<tr>
<td>Self Help Guides</td>
<td>7% 6% 6% 8%</td>
</tr>
<tr>
<td>TOTAL (Count)</td>
<td><strong>429 634 301 39</strong></td>
</tr>
</tbody>
</table>

Table 3: Top main referrals made in semester one

Students were most commonly referred to the Careers and Employer Liaison Centre and the Student Learning Centre in both semesters. This was due mainly to their perception that they...
had chosen the wrong course or their academic preparedness was insufficient to deal with the requirements of the course. These findings are not gender nor age-related, which is consistent with research findings (Yorke, 2010). Referrals to Flinders University Student Association were prominent in both semesters and referrals to Health, Counselling and Disability Services were common only during semester one. Referrals to Admissions became relatively prevalent toward the end of 2014, primarily due to students considering transferring into other Flinders courses.

Similar to Coats’ results (2014), the analysis demonstrates that a substantial number of the students who were successfully contacted had multiple needs that pushed them closer to withdrawing from their studies.

**Conclusions**

There seems to be a general consensus in the literature that student experience in the first year is largely responsible for their level of engagement and subsequent success. The factors involved in this experience need to be continually unpacked to ensure data-based decisions are made to improve student engagement (see James et al., 2010 for a detailed review). As part of this process, the predictive models developed in SPSS Modeller suggest that, unlike the common belief, the relationship between student demographics and their departure intentions were much weaker than initially anticipated. This will need to be explored further before adopting it as a fait accompli.

Although the SSP is still in its infancy, it is a methodical process that complements the transition timeline for first-year students and continues to use existing and new student data to shape its evolution. The findings indicate that the SPSS has already contributed to some at-risk students’ decision to continue with their university studies. The data has also uncovered the level of support needed to enable students to make an informed decision about their future in higher education. In addition, it has informed the various sectors of the University of the ways in which students want and expect to engage with the University. The SSP has created an encouraging level of support from professional and academic staff and, more importantly, from the senior leadership of the University, enabling it to continue and expand beyond 2014.
References


Kift S. (2009). Articulating transition pedagogy to scaffold and to enhance the first year student learning experience in Australian higher education: Final report for ALTC Senior Fellowship Program, Sydney, Australia, ALTC.


Yorke, M. (2010). The quality of the Student Experience: What can institutions learn from data relating to non-completion? *Quality in Higher Education*. 6:1, pp. 61-75; DOI: 10.1080/1358320050001072. [http://dx.doi.org/10.1080/1358320050001072](http://dx.doi.org/10.1080/1358320050001072)