Using Predictive Modelling to Identify Students at Risk of Poor University Outcomes

Pengfei Jia and Tim Maloney

Economics Department, Faculty of Business and Law,
Auckland University of Technology

April 2014

Correspondence to: Tim Maloney, Economics Department, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand, (tim.maloney@aut.ac.nz), 64-9-921-9823.

Abstract

We use predictive modelling to identify students at risk of not successfully completing their first-year courses and not returning to university in the second year. Our aim is two-fold. Firstly, we want to understand the pathways that lead to poor first-year experiences at university. Secondly, we want to develop simple, low-cost tools that would allow universities to identify and intervene on vulnerable students when they first arrive on campus. This is why we base our analysis on administrative data routinely collected as part of the enrolment process from a New Zealand university. We assess the 'target effectiveness' of our model from a number of perspectives. This approach is found to be substantially more predictive than a previously developed risk tool at this university. Students in the top decile of risk scores account for over 29% of first-year course non-completions and more than 23% of second-year student non-retentions at this university.

1. Introduction

Poor outcomes at university are a concern to students, institutions and public funding
bodies. This may be a by-product of the rapidly rising university participation rates
over recent decades in many countries. Course non-completion and dropout rates
may be more common as less able or academically prepared students are admitted to
university. Public funding authorities are also increasingly concerned by the potential
waste of public expenditures on students who subsequently fail at university. For
example, reducing non-completion rates is a core concern of recent reforms of the
tertiary education sector in New Zealand (e.g., see New Zealand Ministry of
Education 2004).

There is a substantial body of empirical literature on the determinants of university
non-completion outcomes (e.g., Wetzel et al. 1999, Montmarquette et al. 2001,
Singell 2004, Kerkvliet and Nowell 2005, Bai and Maloney 2006, Ishitani 2006,
Stratton et al. 2008, and Belloc et al. 2010). Although a comprehensive
understanding of the relative importance of the various reasons for non-completion
behaviour remains elusive, it has been widely recognized that individual
characteristics, student educational backgrounds, and institutional factors are the main
determinants of these outcomes. However, due mainly to limited data availability,
most previous studies have utilized relatively few factors in this analysis. Using a
more comprehensive dataset, our study is able to analyse the impact of a wide variety
of explanatory variables on poor university outcomes.

Our paper uses administrative data from a large public university in New Zealand to
estimate the determinants of course non-completion in the first year and university
non-retention in the second year. Administrative data have a number of advantages
for the purposes of this study. Firstly, these data are gathered as part of the normal
application process, and thus no additional expense or inconvenience is incurred in
acquiring this information. Secondly, because these data are collected for enrolment
purposes, the variables and their definitions are consistent over time. This is an
important aspect if we want to use historical data to predict the at risk status of future
students on an ongoing basis.

This study has two goals. Firstly, we want to estimate the effects of a wide array of factors that may lead to both first-year course non-completion and second-year university non-retention outcomes. Secondly, we want to use these results to test the efficacy of a potential predictive risk tool for the early identification of students who are vulnerable to adverse outcomes at university. This is a trial to show how existing administrative data could be used in targeting intervention services (e.g., special tutorials, student advising and mentoring services) at the most vulnerable students entering university for the first time.

The rest of the paper is organized as follows. Section 2 provides a brief overview of the relevant literature. Section 3 describes the data used in our regression analysis and summarises our econometric approach. Section 4 analyses our empirical results. Finally, Section 5 draws conclusions from this analysis, and suggests possible directions for future work in this area.

2. Literature Review

Predictive risk analysis has been used previously in areas such as health care and child protection (e.g., see Billings et al. 2012, Vaithianathan et al. 2013). To our knowledge, this approach has not been applied previously to the analysis of students at risk of adverse academic outcomes at university.

Yet, there is a substantial literature estimating the factors that influence poor student university experiences. For example, many studies have shown significant effects of student demographic characteristics (e.g., ethnicity, gender, country of origin and age) on dropout behaviour (e.g., see Grayson 1998, Robst et al. 1998, Wetzel et al. 1999, Montmarquette et al. 2001, Bai and Maloney 2006, Mastekaasa and Smeby 2008, Belloc et al. 2010, and Rodgers 2013). Prior academic performance has been found to be related to success at university (e.g., see Betts and Morell 1999, Cohn et al. 2004, Cyrenne and Chan 2012, and Ficano 2012). However, fewer studies have empirically examined the impact of past academic performance on student dropout behaviour (Wetzel et al. 1999, Montmarquette et al. 2001, Singell 2004, Bai and Maloney 2006, Ishitani 2006, Stratton et al. 2008, Belloc et al. 2010, and Ost 2010).

Although there is substantial literature on the effects of class size on academic performance at high school (e.g., see Angrist and Lavy 1999, Krueger 2003, and Rivkin et al. 2005), to our knowledge, no previous published work has considered the effects of class size on academic outcomes at university. We use non-experimental data in our study to estimate the effects of various facets of class size on the probability of course non-completion and university non-retention.

Past research confirms the considerable differences of study areas on student dropout behaviour (e.g., Robst et al. 1998, and Rodgers 2013). Students who study science or engineering may be more likely to drop out than those who study arts or business, possibly due to the degree of difficulty of course material and academic expectations in these programmes.

3. Data and Methodology

Administrative data were provided by a large public university in New Zealand for the purposes of this study. Data were made available on all first-year students who enrolled in Bachelor degree programmes for the first time at this university during the 2009 through 2012 academic years. The full sample contains 18,638 individuals and 101,948 course-specific observations. Individual student observations are used to examine non-retention outcomes in the second year, while individual course observations are used to investigate course non-completion outcomes in the first year.

Variable definitions and descriptive statistics are provided in Table 1. Our dataset contains detailed information typically available at the time of initial enrolment at university (e.g., year of entry, demographic characteristics, high school academic performance, and course and programme enrolment information).

(Insert Table 1 Here)

Two dependent variables are used in this study: course non-completion outcomes in the first-year; and university non-retention outcomes in the second year.[1] The first dummy variable is set equal to one if the student did <u>not</u> successfully complete a course (i.e., receive a passing grade) in the first year; zero otherwise. The second dummy variable is set equal to one if the student did <u>not</u> return to re-enrol at this university at the beginning of the second year; zero otherwise. The results reported in Table 1 show that the mean non-completion rate is 0.154 for the course observations in our sample. The mean non-retention rate is 0.226 for the first-year student observations in our sample. Of course, students may leave university either temporarily or permanently, and for a variety of reasons.[2]

We have data on all first-year students from four annual cohorts. Our observations are fairly evenly distributed across these four years (see Table 1). We include five dummy variables for a student's self-reported ethnicity (i.e., Asian, European, Māori, Pacifica, and other ethnicities). The latter is a residual category of all other reported ethnicities. The final category (Unknown) includes students who did report their ethnicity.

We use seven dummy variables on country of origin. Most first-year students are from New Zealand (69.5%), followed by China (8.6%), Korea (2.2%), India (1.6%) and Vietnam (1.3%). All other reported countries of origin are combined into a residual category 'Others' accounting for 15.5% of first-year students. Those not reporting their country of origin make up 1.3% of our sample. Other personal characteristics include being female (60.1% of our sample) and enrolling for study part-time (29.3%). Just over one-half of our first-year students (57.8%) report information on their first or primary language. Of those who do, 70.7% identify English as their first language. Domestic students are defined as those receiving

---

[1] We do not distinguish in this analysis between course dropouts (i.e., individuals who discontinued study prior to the end of the semester and dropped out of the course) and course failures (i.e., individuals who continued to the end of the semester, completed all assessments, but failed the course). This is largely because of the government reporting requirements in New Zealand that emphasise non-completion outcomes as a result of either process.

[2] Possible explanations for dropout behaviour include students struggling academically at university, transferring to other institutions, leaving for employment opportunities, etc. It should be noted that we have no information in the database on the reasons why individuals may have failed to return to this university at the beginning of the second year.

domestic funding status (i.e., government subsidies). They comprise 87.9% of the first-year students at this university. The mean age for first-year students is 22.075.

Our dataset contains some information on the high school records of these students. Most high school students in New Zealand sit the National Certificate of Educational Achievement (NCEA) exams in the last three years at school.[3] These are national end-of-the-year exams across a number of compulsory and optional subject areas. Our dataset includes a summary measure of the overall performance on these NCEA exams in the final year of high school. As indicated in Table 1, this NCEA score is available for less than half of the first-year students across our cohorts ('Known NCEA Score'). We explain below how students can enter this university without NCEA results.

Two additional variables are available on the educational background of our students. A value of one for the variable 'Literacy/Numeracy' indicates that tests was taken during high school to investigate possible issues over appropriate literacy and numeracy levels. In New Zealand, high schools are sorted into deciles based on the socio-economic status of residents in the school catchment area.[4] For example, a decile 1 high school is among the 10% of schools from poorest socio-economic areas, while a decile 10 high school is from the wealthiest socio-economic areas. The mean school decile in our sample is 6.846, indicating that these first-year students are drawn predominantly from higher decile schools.

There are six specified ways in which these first-year students could be granted entry to this university. The most conventional entrance type is 'NCEA Admission'. More than one-third of first-year students in our sample (36.3%) gain admission to this university through their NCEA scores. Other students enter through 'Special Admission' status which refers to entering students who did not meet the NCEA entrance requirements, but entered because of their other experiences (i.e., they had reached age 21 or above). Special Admissions account for 13.0% of first-year

---

[3] For more information on the NCEA system see http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/.

[4] For more information on the process used to determine school deciles see http://www.minedu.govt.nz/NZEducation/EducationPolicies/Schools/SchoolOperations/Resourcing/OperationalFunding/Deciles/HowTheDecileIsCalculated.aspx.

students in our sample. The variables 'Internal' and 'External' refer to students who gained university entrance because of previous study at this or another university. Internal entrants (8.9%) had completed a 'pre-degree' certificate or diploma at this particular university. External entrants (15.0%) had previously attended another university. A few students enter this university through the completion of Cambridge or International Baccalaureate programmes at secondary school ('Cambridge/IB'). Relatively few students enter through these more prestigious and challenging programmes (1.4%). Finally, there are a number of other ways in which students can gain entry to this university. These are included in the residual category 'Others', and comprise slightly more than one-quarter of all first-year students in our sample (25.3%). Primarily, these include foreign students who gain entry through equivalent overseas high school qualifications.

For the purpose of analysing course non-completion outcomes, our administrative dataset contains some potentially useful information on the characteristics of these courses. From the recorded information at the outset of the academic year, we know the recommended 'Study Hours' in a course. Most courses (84.4%) report 'Known Contact' hours. These include scheduled lecture, tutorial, workshop and lab hours. They could also include scheduled office hours, and group study hours and generic academic preparation workshops in areas such as English, writing skills and mathematics. We also know the average 'Class Size' and the 'Course Size'. The latter is the total number of students enrolled in the course. The former is the average number of students in a classroom. For example, a large first-year course could have 1,000 students enrolled. This would be the course size. These students could be taught in a single large class of 1,000, or they could be taught in 20 classes with an average class size of 50 students. We consider the separate effects of both course and class size on course non-completion outcomes. We also know whether or not the course is supported with internet content. Finally, we know the academic level of the course. A 'Level 4' course contains content that is intended for students below a Baccalaureate degree level. Most courses taken by these first-year students (83.6%) are intended for the first year of university study ('Level 5'), but some students enrol

in courses intended for second and third-year study ('Level 6' (15.6%) and 'Level 7' (0.4%), respectively).[5]

We have additional academic information on students including the number of courses in which they have enrolled ('Courses Taken') and the proportion taken at Levels 6 or 7 ('High Level'). We also know whether or not the student has enrolled for a double-degree programme, and whether or not study is relegated to a single campus at this university. Less than 1% of first-year students enrol for a double degree, partly due to stringency of the entry requirements.

Finally, our dataset contains information on the initial programme of study. We use dummy variables to identify the largest 11 Bachelor degree programmes. The residual category includes all of the smaller degree programmes (7.5% of students in our sample). The largest three programmes are the Bachelor of Business (28.2%), the Bachelor of Health Sciences (19.5%), and the Bachelor of Arts (7.8%).

Maximum likelihood probit analysis will be used to estimate the effects of these various individual, school and enrolment factors on our two dummy dependent variables. The basic probit model can be written:

$$Y_i^* = \boldsymbol{\beta X_i} + u_i \tag{1}$$

where Y_i^* is a latent variable associated with course non-completion or student non-retention propensities. What we observe is a dummy variable Y_i that equals 1 if the course was not successfully completed in the first year, or the student did not return to re-enrol at this university in the second year; 0 otherwise. This depends on the latent dependent variable crossing an arbitrary threshold of zero.

$$Y_i = \begin{cases} 1 & if \ Y_i^* > 0 \\ 0 & if \ Y_i^* \leq 0 \end{cases} \tag{2}$$

---

[5] The typical university baccalaureate programme in New Zealand is completed in three years of full-time study.

All of the aforementioned factors are included in the vector $\boldsymbol{X_i}$. The unknown coefficients are represented by the $\boldsymbol{\beta}$ vector which will need to be estimated. The probability of course non-completion or student non-retention can be denoted as the following:

$$P(Y_i^* > 0) = P(\boldsymbol{\beta X_i} + u_i > 0) = P(u_i > -\boldsymbol{\beta X_i}) = \boldsymbol{\Phi}(\boldsymbol{\beta X_i}) \tag{3}$$

where $\boldsymbol{\Phi}(\cdot)$ is the Cumulative Distribution Function (CDF) of the standard normal.

We will use the average marginal effects to describe the influence of a one-unit change in an explanatory variable on the probability of course non-completion or student non-retention. This is because a Probit model is a non-linear function of the coefficients, and the marginal effects are dependent on the values of the other regressors. For any particular factor X_k, this partial derivative can be written:

$$\frac{\partial P(Y_i = 1|\boldsymbol{X_i})}{\partial X_k} = \beta_k \phi(\boldsymbol{\beta X_i}) \tag{4}$$

where $\phi(\cdot)$ is the Probability Distribution Function (PDF) of the standard normal.[6]

This probit estimation is also used in the development of our Predictive Risk Models (PRMs), which can be used to generate risk scores for any first-year student enrolling at this university. To assess the effectiveness of these predictive risk tools, we compare our predicted outcomes against the actual outcomes for each of our dependent variables. For this reason, our full sample is randomly split into two equal-sized 'estimation' and 'validation' samples (e.g., see medical applications of this methodology in Billings et al. 2012). The estimation samples will be used to estimate the probit models, and the validation samples will be used to assess how well the PRMs correctly identify the actual course non-completion outcomes and student non-retention outcomes.

---

[6] The marginal effects could also be calculated at the sample means for the explanatory variables. For continuous functions in large samples, this technique yields similar results to the sample mean for the individual marginal effects.

4. Empirical Results

Two separate maximum likelihood probit models were estimated for this study, using the estimation samples for course non-completion outcomes in the first year and student non-retention outcomes in the second year. The estimated coefficients, standard errors, and average marginal effects are presented in Table 2 for both dependent variables.

(Insert Table 2 Here)

4.1 Results on First-Year Course Non-Completion

Holding constant other measured factors, we find that the probability of course non-completion varies systematically across the years (where 2012 is the excluded or benchmark year). All three estimated coefficients on the included year dummies are negative and statistically different from zero at better than a 1% level. This says that the course non-completion probability was highest in 2012 compared to the previous three years. However, given the lack of any clear time trend in these estimated marginal effects, it would be premature to conclude that these results suggest a systematic increase in course non-completion rates over time.

Ethnicity appears to have a substantial impact on the probability of course non-completion. Relative to the omitted category ('Unknown' ethnicity), being either European or Asian has a statistically significant negative impact on the probability of course non-completion. The estimated partial derivatives indicate that these mean reductions in the probability of course non-completion are -3.76 percentage points for Europeans and -2.60 percentage points for Asians. Being either Pacifica or Māori has a statistically significant positive impact on course non-completion. This suggests that the difference between the estimated probabilities of course non-completion for an otherwise observationally equivalent first-year Pacifica student is approximately 10.43 percentage points higher than that of a European student.

The results on country of origin require some explanation. The estimated coefficients on all six dummy variables are negative and statistically significant at better than a 1% level, compared to those who did <u>not</u> report their country of origin. In other words, those not reporting a country of origin at the time of initial enrolment appear to be the highest risk group.

Consistent with earlier studies in this literature, female students have a relatively lower estimated probability of course non-completion. Holding other things constant, being female lowers this probability of course non-completion by 2.73 percentage points. This effect is statistically significant at better than 1% level.

Studying part-time study is estimated to substantially increase the rate of course non-completion. Being a part-time student increases this probability by 15.49 percentage points. English as the first language has no measurable impact on course non-completion rates in our sample. Being a domestic student increases the probability of course non-completion by 3.80 percentage points.

We use a series of dummy variables to allow for flexibility in the age effects on course non-completion outcomes. One dummy variable is used for being under the age of 18. A series of eight dummies are used for individual ages from 18 through 25 inclusive, and three dummies are used for age ranges 26 through 30, 31 through 35 and 36 through 45. The omitted age group is 46 and older. All individual ages from 18 through 25 have positive and significant effects on the probability of course non-completion. The results suggest that first-year students aged 20 or 21 are at the highest risk. Their probabilities of course non-completion are estimated to be, respectively, 7.13 and 7.02 percentage points higher than those of students aged 45 or older.

Students who scored higher on their NCEA exams are found to be at lower risk of course non-completion during their first year at university. Two variables must be considered in interpreting these results. The first is a dummy variable on having information on these exam results ('Known NCEA Score'), and the second is the composite exam score from this last year of high school ('Actual NCEA Score'). The

estimated effect of having an NCEA score on this probability of course completion could be written:

$$\frac{\partial P(Non - Completion)}{\partial Known\ NCEA\ Score} = 7.60\% - 0.10\% \cdot Actual\ NCEA\ Score \qquad (5)$$

We know from Table 1 that the sample mean for those reporting a NCEA score is 155.107. Thus, for the average student with NCEA results, these exams reduce the probability of course non-completion in the first year by an average of nearly 8 percentage points:

$$\frac{\partial P(Non - Completion)}{\partial Known\ NCEA\ Score} = 7.60\% - 0.10\% \cdot 155.107 \approx -7.911\% \qquad (6)$$

The previous section indicated that the dummy variable on 'Literacy/Numeracy' tests in high school picks up possible concerns over the reading, writing and mathematics skills for students. As expected, taking these tests is associated with a significant increase of 1.68 percentage points in the average probability of course non-completion.

We expected that students from lower school deciles would have higher probabilities of paper non-completion during the first year at university. This result is largely confirmed by our analysis, but some discussion around these findings is needed. Firstly, the omitted category includes mostly overseas students who did not come from a high school with a decile ranking. All ten of the school deciles have positive and statistically significant effects on the probability of course non-completion in the first year at university. This again indicates that domestic students are generally at higher risk compared to international students. Secondly, although the largest estimated effects are found for the bottom four deciles, there is no evidence in these results that students from the top decile schools are at the lowest risk of course non-completion. In fact, there is some evidence of a 'U-shaped' relationship. The estimated effects for deciles 8 through 10 are positive and larger in magnitude than deciles 5 through 7. One possible explanation for these results is that many first-year students at this university coming from the highest decile schools were unable to gain

admittance to higher ranked universities in New Zealand and overseas and generally did not have the same academic preparation (or motivation) of students from mid-decile schools.

The entrance types for students have measurable impacts on the probability of course non-completion in the first year. Students entering with a Cambridge or International Baccalaureate qualification are at the lowest risk. This entrance type reduces the probability of course non-completion by an average of 6.44 percentage points. Students with an 'External' entry (i.e., previous study at another university) are next lowest risk group, while 'Internal' entry (i.e., holding a pre-degree qualification from this university) has no measurable impact on course non-completion. The two highest risk entrance types are 'Special Admission' and 'NCEA Admission'. The NCEA Admission standard is closely connected to the effect of NCEA Score discussed earlier, and isn't likely to be a risk indicator because the reported NCEA scores for these students reduce the probability of course non-completion. However, the Special Admission standard is likely to be an indicator of vulnerability, because these are generally older students who lack school qualifications and enter on the basis of their age and work experience. This Special Admissions effect combined with at risk nature of students aged in their early twenties makes this a particularly vulnerable group.

The remaining covariates in this regression model relate to the courses or degree programmes in which these students were enrolled during their first year at university. As mentioned previously, we draw a distinction between the overall number of students enrolled in a course ('Course Size') and the average number of students in a classroom ('Class Size'). To ease the interpretation of the estimated results, both variables are divided by 10. Individuals often enrol in large first-year courses, but these can be taught in either large settings (e.g., a single mass lecture) or small settings (e.g., multiple streams taught in smaller classrooms). These course and class size effects could be quite different for the probability of course non-completion. For example, courses with large enrolments could reduce the probability of course non-completion because of the introductory nature of the subject material and the need for large-scale assessments. On the other hand, similar to the usual justification in the literature on class size effects at school, large classroom settings could increase the

probability of course non-completion due to the lack of individual attention for students. These are precisely the direction of the effects that we find in our analysis. The estimated course size effect is negative and statistically significant at better than a 1% level. We find that a course enrolment increase of 10 students, on average, reduces the probability of course non-completion by 0.03 percentage points. The estimated class size effect is positive and statistically significant at better than a 5% level. We find that a classroom size increase of 10 students, on average, raises the probability of course non-completion by 0.10 percentage points. This suggests that course non-completion rates would be lowered by enrolling students in large first-year courses, but teaching them in smaller classroom settings.

Finally, our results indicate that programme study areas play an important role in course non-completion outcomes in the first year. We know the degree programmes in which these students initially enrolled, which includes multiple programmes for those doing a double degree. Relative to the reference group of students in relatively small programmes, three distinct programmes had significantly higher estimated rates of course non-completion. They are, in order of the size of these positive marginal effects, Bachelor of Mathematical Science (BMS 3.62 percentage points), Bachelor of Engineering Technology (MEngT 1.84 percentage points), and Bachelor of Arts (BA 1.72 percentage points). Seven programmes had significantly lower estimated rates of course non-completion. In order of the size of these negative marginal effects, the largest three programmes were Bachelor of Education (BEdu -10.89 percentage points), Bachelor of Design (BDe -6.79 percentage points), and Bachelor of Communication Studies (BCS -6.10 percentage points). Some caution should be exercised in interpreting these results. They could indicate something about the rigour or difficulty of first-year study in these areas, but could equally indicate something about the unobserved characteristics of the students who enrol in these degree programmes.

4.2 Results on Second-Year University Non-Retention

The last three columns of Table 2 report the regression results on the non-retention outcomes in the second year for our estimation sample of students. Recall that both Pacifica and Māori students were significantly more likely to not complete their first-

year courses compared to other ethnic groups. The only ethnic group with a statistically significant effect on non-retention is Māori. Specifically, Māori students have a probability of non-retention that is, on average, 5.85 percentage points higher than students without a reported ethnicity. This suggests that Pacifica students are the most likely ethnic group to not complete their courses in the first year, while Māori students are the most likely ethnic group to not return to the university in the second year.

The estimated results on the country of origin variables have similar negative and significant effects for both student non-retention and course non-completion. This suggests that students <u>not</u> reporting a country of origin are in the highest risk group for both adverse outcomes. Female students are at lower risk of both course non-completion and non-retention. However, the latter effect is smaller in magnitude and weaker in statistical significance. Part-time students are substantially more likely to drop out of university in the second year. Studying part-time increases the probability of non-retention by an average of 15.80 percentage points. Thus, part-time study is arguably the single most important single at risk factor for poor university outcomes. Domestic students are relatively more likely to discontinue their study at this university in the second year. Although age seemed to play an important role in course non-completion outcomes in the first year, it has no measurable effect on the probability of non-retention in the second year.

Students who scored higher on their NCEA exams in high school are less likely to drop out of university in the second year. Again, we need to estimate this overall impact using the two estimated average marginal effects:

$$\frac{\partial P(Non-Retention)}{\partial Known\ NCEA\ Score} = 5.30\% - 0.06\% \cdot Actual\ NCEA\ Score \qquad (7)$$

Using the sample means from Table 1, we estimate that for those reporting these exam results, they reduce the probability of student non-retention by over 4 percentage points.

$$\frac{\partial P(Non - Retention)}{\partial Known\ NCEA\ Score} = 5.30\% - 0.06\% \cdot 155.107 \approx -4.006\% \qquad (8)$$

We find that 'Literacy/Numeracy' tests in high school are associated with a significant increase of 3.74 percentage points in the probability of non-retention in the second year. Recall that students reporting a school decile were more likely to <u>not</u> complete their first-year courses. It is noteworthy that the same result does <u>not</u> hold for non-retention. None of the estimated coefficients on school deciles are positive and significant. In fact, students coming from schools in declies 4, 6, 7 and 10 are significantly less likely to be university dropouts in the second year.

Students entering this university with a Cambridge or International Baccalaureate qualification are both more likely to complete their first-year courses and to be retained by the university in the second year. Both 'External' and 'Internal' entry reduce the probability of non-retention in the second year. 'Special Admission' status, which was a risk factor for course non-completion, has no measurable effect on student non-retention.

We found earlier that students studying part-time are one of the most vulnerable groups for both course non-completion and university non-retention. In a similar way, enrolling in 6 or more courses and in a double degree both substantially reduce the probability of non-retention in the second year.

Finally, of the three highest-risk programmes for course non-completion, only the Bachelor of Arts degree has a significantly positive estimated effect on student non-retention. Of the three lowest-risk programmes for course non-completion, two of them have significantly negative estimated coefficients on non-retention. The degree programme with the lowest risk of course non-completion also has the lowest risk of university non-retention (the Bachelor of Education).

4.3 Assessing the Predictive Power of Our PRMs

One way to assess the overall performance of our probit regression models is to consider the Pseudo R^2 statistics reported at the bottom of Table 2. The usual

interpretation is that our models can explain approximately 13.19% of the variation in course non-completion outcomes in the first year and 10.63% of the variation in university non-retention outcomes in the second year, respectively. These statistics, of course, only summarise the predictive power of our analysis *within* these estimation samples. We want to know how well these models perform in predicting these outcomes *outside* of these samples.

We report the area under the Receiver Operator Characteristic (ROC) curves for both course non-completion and student non-retention outcomes in the summary statistics of Table 2 using these respective validation samples. The ROC curves characterise the relationship between the 'sensitivity' and 'specificity' in these two models. Sensitivity is the probability that a course failure (or student dropout) outcomes is correctly identified. Specificity is the probability that a course completion (or student retention) outcome is correctly identified. We graphically illustrate the trade-offs between sensitivity and one minus the specificity at all possible thresholds. These results are shown in Figures 1 and 2.

(Insert Figures 1 and 2 Here)

The area under the ROC curve for course non-completion is 0.7553. This indicates that there is a 75.53% probability that a randomly selected course observation with a non-completion outcome will receive a higher risk score from our Predictive Risk Model (PRM) than a randomly selected course observation with a completion outcome. This is an indicator of the 'target effectiveness' of this predictive risk tool could be compared to the results from other types of analyses. Similar interpretations can be given for the non-retention analysis with the area under ROC curve at 0.7125.

Another approach is assessing the effectiveness of our PRMs is to compare predicted to actual outcomes. We use the regression results reported in Table 2 to compute risk scores for all course non-completion and student non-retention outcomes in our validation samples. We then rank these predicted probabilities and sort them into deciles, and determine the proportion of actual adverse outcomes that would be captured at every decile. Suppose we wanted to intervene (i.e., provide specific services) to students in the top decile (i.e., those with the highest 10% of risk scores).

If our models were completely ineffective at predicting these outcomes, then the top 10% of risk scores would account for only 10% of the actual adverse outcomes. The results in Table 3 indicate that the highest 10% of risk scores in the validation samples would capture 29.25% of actual course non-completions in the first year and 23.33% of actual student non-retentions in the second year. If we targeted the top two deciles, we would capture 47.57% of course non-completions and 40.91% of student non-retentions.

(Insert Table 3 Here)

It's often difficult to provide any meaningful relative comparisons to the predictive power analysis of a particular PRM. Fortunately, in this situation we had information on an existing risk analysis tool developed by this university, which provides a convenient benchmark. The university had previously used the results from a survey administered to first-year students to predict who would likely experience academic difficulties over the first year of study. University administrators attached 'weights' to the survey responses based on subjective assessments on the relative importance of these various factors, and not on a formal statistical analysis of the relationships between these variables and course non-completion outcomes.

We constructed the risk scores from our validation sample using this existing administrative tool, and compared these predicted outcomes to observed course non-completions.   By any measure, the predictive power of this administrative tool was substantially inferior to our PRM. Because of 'ties' in adding up these risk measures using the integer weights, we can't select only the highest 10% of risk scores. The approximate 'top decile' using the university's administrative tool accounted for 11.78% of course outcomes, and these captured 23.51% of actual course non-completions in our validation sample. The top decile of risk scores using our PRM was nearly three-times more likely to capture a course non-completion than the overall sample (29.25/10.00). The top decile of risk scores using the administrative tool was less than two-times more likely to experience a course non-completion (23.51/11.78). In this sense, our PRM was 46.5% more 'target effective' than the existing administrative tool.

The same comparisons can be made for the top two deciles. Again, because of ties, the existing administrative tool accounted for 25.27% of course outcomes, but captured only 39.11% of actual course non-completions. The top two deciles of risk scores using our PRM was 2.4-times more likely to experience a course non-completion than the overall sample (47.57/20.00). The top two deciles of risk scores using the administrative tool was 1.5-times more likely to experience a course non-completion (39.11/25.27). In this sense, the 'hit rate' of our PRM is approximately 53.4% higher than the existing administrative tool.

This relatively better performance of our PRM is not that surprising given that the administrative tool used by the university had never been appropriately validated. This PRM approach has a very important additional advantage. The survey-based administrative tool requires the dissemination and processing of a first-year student survey each year. This can be an expensive operation. Our PRM tool is based entirely on routine data collected as part of the enrolment process. Thus, once developed, there is virtually no additional on-going cost in using this PRM approach. In this sense, it is relatively more 'target effective' and 'cost efficient'.

5. Conclusion

This study has empirically estimated the determinants of course non-completion outcomes in the first year and student non-retention outcomes in the second year using administrative data from a large public university in New Zealand. These Predictive Risk Models (PRMs) have been developed to improve our understanding of the factors that that place students at risk of adverse outcomes early in their university careers. In addition, these PRMs could be used by universities to develop effective, low-cost tools for identifying students at risk of adverse outcomes, and to provide early interventions for students that are most likely struggle at university.

The two dependent variables used in our regression analysis were course non-completion outcomes in the first year and student non-retention outcomes in the second year. Administrative data were taken from four annual cohorts of students entering degree programmes for the first time at this university. Our findings suggest

that a wide array of factors influence the probabilities of course non-completion and student non-retention. For example, part-time study is estimated to substantially raise the probabilities of both detrimental outcomes. Pacifica students are the ethnic group most at risk of course non-completion outcomes in the first year, while Māori students are the ethnic group most at risk of non-retention outcomes in the second year. Females are at lower risk of both course non-completions and student non-retention outcomes. Better results on national high school exams substantially reduce the risk of both adverse outcomes. Larger overall course enrolments, but smaller average class sizes are associated better course outcomes. Finally, early university experiences vary substantially across the degree programmes.

The areas under ROC curves were 0.7553 and 0.7125, respectively, for the course non-completion and student non-retention outcomes. The top risk decile of course observations identified by PRM could account for 29.25% of actual course non-completion outcomes. The top risk decile of student observations could account for 23.33% of actual student non-retention outcomes. These results were superior to the existing administrative tool used by this university. Our PRM is at least 46.5% more target effective in identifying students vulnerable for course non-completions. We also claim that out PRM would also be relatively more cost-effective because it would be based on existing administrative data already collected as part of the enrolment process.

There is more that can be done in this area to better understand the determinants of these early adverse outcomes at university, and to improve the accuracy of any PRM for identifying at risk students. We could improve our measures of these early university outcomes. For example, we've concentrated on the non-completion outcomes for courses. This doesn't distinguish between students who discontinue their study early in the semester (i.e., course dropouts) and those who don't meet the passing standards at the end of the semester (i.e., course failures). More could be done to expand the range of covariates used in the regression analysis. For example, we have no information in our administrative data on parental education, family finances, student scholarships or other financial aid, and peer and community characteristics. We could also do more with existing administrative data to improve the quality of our predictive variables. For example, we have access to only partial

information on student academic performance in high school. It would be possible with available data from the Ministry of Education to gain access to the results from national exams for students over their two previous years at high school. This could greatly improve the quality of our predictive risk tool, and again help the university in targeting its limited resources at the most vulnerable students.

References

Angrist, J., & Lavy, V. (1999). Using Maimonides's Rule to estimate the effect of class size on children's academic achievement. Quarterly Journal of Economics, 114, 533-575.

Bai, J., & Maloney, T. (2006). Ethnicity and academic success at university. New Zealand Economic papers, 40 (2), 181-213.

Belloc, F., Maruotti, A., & Petrella, L. (2010). University drop-out: an Italian experience. Higher Education, 60, 127-138.

Betts, J. R., & Morell, D. (1999). The determinants of undergraduate grade point average: the relative importance of family background, high school resources, and peer group effects. Journal of Human Resources, 34 (2), 268-293.

Billings, J., Blunt, I., Steventon, A., Georghiou, T., Lewis, G., & Bardsley, M. (2012). Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30). BMJ Open, 2 (4), e001667.

Cohn, E., Cohn, S., Balch, D. C., & Bradley, J. (2004). Determinants of undergraduate GPAs: SAT scores, high-school GPA and high school rank. Economics of Education Review, 23, 577-586.

Cyrenne, P., & Chan, A. (2012). High school grades and university performance: a case study. Economics of Education Review, 31, 524-542.

Ficano, C. C. (2012). Peer effects in college academic outcomes – gender matters! Economics of Education Review, 31, 1102-1115.

Grayson, J. P. (1998). Racial origin and student retention in a Canadian university. Higher Education, 36, 323-352.

Ishitani, T. T. (2006). Studying attrition and degree completion behaviour among first-generation college students in the United States. Journal of Higher Education, 77 (5), 861-885.

Kerkvliet, J., & Nowell, C. (2005). Does one size fit all? University differences in the influence of wages, financial aid and integration on student retention. Economics of Education Review, 24, 85-95.

Krueger, A. B. (2003). Economic considerations and class size. Economic Journal, 113, F34-F63.

Mastekaasa, A., & Smeby, J. C. (2008). Educational choice and persistence in male- and female-dominated fields. Higher Education, 55, 189-202.

Montmarquette, C., Mahseredjian, S., & Houle, R. (2001). The determinants of university dropouts: a bivariate probability model with sample selection. Economics of Education Review, 20, 475-484.

New Zealand Ministry of Education. (2004). Retention, completion and progression in tertiary education 2003. Wellington: Ministry of Education.

Ost, B. (2010). The role of peers and grades in determining major persistence in the sciences. Economics of Education Review, 29 (6), 923-934.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools and academic achievement. Econometrica, 73 (2), 417-458.

Robst, J., Keil, J., & Russo, D. (1998). The effect of gender composition of faculty on student retention. Economics of Education Review, 17 (4), 429-439.

Rodgers, T. (2013). Should high non-completion rates amongst ethnic minority students be seen as an ethnicity issue? Evidence from a case study of a student cohort from a British university. Higher Education, 66 (5), 535-550.

Singell, L. D. (2004). Come and stay a while: does financial aid affect retention conditioned on enrolment at a large public university? Economics of Education Review, 23, 459-471.

Stratton, L. S., O'Toole, D. M., & Wetzel, J. N. (2008). A multinomial logit model of college stopout and dropout behaviour. Economics of Education Review, 27, 319-331.

Tinto, V. (1993). Leaving College: Rethinking the Causes and Cures of Student Attrition, 2nd Edition. Chicago: Chicago University Press.

Vaithianathan, R., Maloney, T., Putnam-Hornstein, E., & Jiang, N. (2013). Using predictive modelling to identify children in the public benefit system at high risk of substantiated maltreatment. American Journal of Preventive Medicine, 45 (3), 354-359.

Wetzel, J. N., O'Toole, D. M., & Peterson, S. (1999). Factors affecting student retention probabilities: a case study. Journal of Economics and Finance, 23 (1), 45-55.

Table 1
Descriptive Statistics and Variable Definitions
Full Sample

Variable	Definition	Mean (std. deviation)
Dependent Variables		
Non-Completion	1 if first-year course is completed; zero otherwise	0.154 (0.361)
Non-Retention	1 if student returns to university in the second year; zero otherwise	0.226 (0.418)
Year of Cohort		
Year 2009	1 if student first enrols in the year 2009; zero otherwise	0.225 (0.478)
Year 2010	1 if student first enrols in the year 2010; zero otherwise	0.253 (0.435)
Year 2011	1 if student first enrols in the year 2011; zero otherwise	0.240 (0.427)
Year 2012	1 if student first enrols in the year 2012; zero otherwise	0.281 (0.450)
Ethnicity		
Asian	1 if student reports ethnicity as Asian; zero otherwise	0.242 (0.428)
European	1 if student reports ethnicity as European; zero otherwise	0.392 (0.488)
Māori	1 if student reports ethnicity as Māori; zero otherwise	0.098 (0.297)
Pacifica	1 if student reports ethnicity as Pacifica; zero otherwise	0.112 (0.316)
Others	1 if student reports other ethnicity; zero otherwise	0.080 (0.272)
Unknown	1 if students reports no ethnicity; zero otherwise	0.076 (0.265)
Country of Origin		
New Zealand	1 if student reports New Zealand as country of origin; zero otherwise	0.695 (0.460)
China	1 if student reports China as country of origin; zero otherwise	0.086 (0.280)
India	1 if student reports India as country of origin; zero otherwise	0.016 (0.127)
Korea	1 if student reports Korea as country of origin; zero otherwise	0.022 (0.147)
Vietnam	1 if student reports Vietnam as country of origin; zero otherwise	0.013 (0.113)
Others	1 if student reports other country of origin; zero otherwise	0.155 (0.362)
Unknown	1 if students reports no country of origin; zero otherwise	0.013 (0.111)
Personal Characteristics		
Female	1 if student is female; zero if male	0.601 (0.490)
Part-Time	1 if student is enrolled part-time; zero full-time	0.293 (0.455)
Language	1 if student reports a first language; zero otherwise	0.578 (0.494)
English	1 if student reports English as first language; zero otherwise (conditional on reporting a first language)	0.707 (0.455)
Domestic	1 if student receives domestic funding; zero otherwise	0.879 (0.326)
Age	Mean age	22.075 (6.322)
High School Information		
Known NCEA Score	1 if NCEA score is available from last year of school; zero otherwise	0.444 (0.497)
Actual NCEA Score	Actual NCEA score (conditional on availability of score)	155.107 (62.860)
Literacy/Numeracy	1 if student took literacy and numeracy test in school; zero otherwise	0.238 (0.426)
School Decile	Mean school decile (conditional on availability of school decile)	6.846 (2.812)

Entrance Type

NCEA Admission	1 if student entered through NCEA level 3; zero otherwise	0.363 (0.481)
Special Admission	1 if student entered through Special Admission category; zero otherwise	0.130 (0.336)
Internal	1 if student entered through pre-degree at this University; zero otherwise	0.089 (0.285)
External	1 if student entered through study at another university; zero otherwise	0.150 (0.358)
Cambridge/IB	1 if student entered through Cambridge or International Bachelaureate; zero otherwise	0.014 (0.117)
Others	1 if student entered through some other category; zero otherwise	0.253 (0.435)

Course Information

Study Hours	Recommended hours of class and preparation time over the semester	180.539 (62.964)
Known Contact	1 if contact hours for the paper are reported	0.844 (0.362)
Contact Hours	Contact hours for the paper (conditional on reporting contact hours)	75.980 (32.234)
Class Size	Average class size in the course	38.279(28.932)
Course Size	Total number of students enrolled in the course	562.194 (535.464)
Internet Content	1 if course is supported with internet content; zero otherwise	0.588 (0.492)
Level 4	1 if course is at level 4 (pre-degree); zero otherwise	0.005 (0.067)
Level 5	1 if course is at level 5 (first year); zero otherwise	0.836 (0.371)
Level 6	1 if course is at level 6 (second year): zero otherwise	0.156 (0.363)
Level 7	1 if course is at level 7 (third year); zero otherwise	0.004 (0.059)

Individual Academic Information

Number of Courses	Number of courses taken by the student	5.243 (2.301)
High Level	Proportion of level 6 or 7 courses taken by the student	0.138 (0.202)
Double Degree	1 if student is enrolled in a double-degree; zero otherwise	0.008 (0.088)
One Campus	1 if student is taking all courses on a single campus; zero otherwise	0.913 (0.283)

First-Year Programmes of Entry

BA	1 if student enrolled in Bachelor of Arts; zero otherwise	0.078 (0.268)
BBus	1 if student enrolled in Bachelor of Business; zero otherwise	0.282 (0.450)
BCIS	1 if student enrolled in Bachelor of Computer Information Science; zero otherwise	0.049 (0.216)
BCS	1 if student enrolled in Bachelor of Communication Studies; zero otherwise	0.068 (0.250)
Bde	1 if student enrolled in Bachelor of Design	0.074 (0.262)
BEdu	1 if student enrolled in Bachelor of Education	0.040 (0.197)
BEngT	1 if student enrolled in Bachelor of Engineering Technology; zero otherwise	0.029 (0.168)
BHS	1 if student enrolled in Bachelor of Health Science; zero otherwise	0.195 (0.396)
BIHM	1 if student enrolled in Bachelor of International Hospitality Management; zero otherwise	0.043 (0.204)
BMS	1 if student enrolled in Bachelor of Mathematical Science; zero otherwise	0.006 (0.079)
BSR	1 if student enrolled in Bachelor of Sports and Recreation; zero otherwise	0.060 (0.237)
Others	1 if student enrolled in another smaller programme; zero otherwise	0.075 (0.291)

	Table 2 **Estimated Results from Maximum Likelihood Probit Analysis on** **Course Non-Completion and Student Non-Retention** **Estimation Subsample**					

	Course Non-Completion in First Year			**Student Non-Retention in Second Year**		
Variable	**Coefficient**	**Std. Error**	**dy/dx**	**Coefficient**	**Std. Error**	**dy/dx**
Constant	-0.6693***	0.1148	-	-0.1374	0.2156	-
Year of Cohort						
Year 2009	-0.1069***	0.0217	-2.20%	-0.0134	0.0447	-0.36%
Year 2010	-0.0752***	0.0198	-1.54%	0.1030**	0.0424	2.74%
Year 2011	-0.1725***	0.0202	-3.54%	0.1112***	0.0428	2.96%
Ethnicity						
Asian	-0.1267***	0.0442	-2.60%	-0.1079	0.0901	-2.87%
European	-0.1830***	0.0483	-3.76%	0.0164	0.0984	0.44%
Māori	0.1666***	0.0515	3.42%	0.2200**	0.1064	5.85%
Pacifica	0.3247***	0.0501	6.67%	0.0967	0.1040	2.58%
Others	-0.0417	0.0511	-0.86%	-0.0525	0.1048	-1.40%
Country of Origin						
New Zealand	-0.4828***	0.0630	-9.91%	-0.7431***	0.1258	-19.78%
China	-0.4488***	0.0714	-9.21%	-0.8090***	0.1411	-21.53%
India	-0.4670***	0.0865	-9.59%	-0.7482***	0.1777	-19.91%
Korea	-0.3356***	0.0809	-6.89%	-0.4329***	0.1616	-11.52%
Vietnam	-0.6662***	0.1081	-13.68%	-1.5075***	0.2574	-40.12%
Others	-0.4833***	0.0650	-9.92%	-0.8868***	0.1299	-23.60%
Personal Characteristics						
Female	-0.1328***	0.0165	-2.73%	-0.0583*	0.0343	-1.55%
Part-Time	0.7546***	0.0179	15.49%	0.5935***	0.0466	15.80%
Language	0.0302	0.0250	0.62%	0.0304	0.0523	0.81%
English	0.0057	0.0269	0.12%	-0.0144	0.0568	-0.38%
Domestic	0.1852***	0.0434	3.80%	0.2495***	0.0880	6.64%
Under Age 18	0.1244	0.1098	2.55%	-0.2035	0.2180	-5.42%
Age 18	0.1854***	0.0645	3.81%	-0.1700	0.1243	-4.53%
Age 19	0.2233***	0.0646	4.58%	-0.0735	0.1245	-1.96%
Age 20	0.3473***	0.0649	7.13%	-0.0037	0.1250	-0.10%
Age 21	0.3418***	0.0658	7.02%	0.0469	0.1272	1.25%
Age 22	0.2411***	0.0679	4.95%	-0.0147	0.1323	-0.39%
Age 23	0.2555***	0.0695	5.25%	-0.082	0.1353	-2.18%

Age 24	0.1512**	0.0730	3.10%	-0.0653	0.1421	-1.74%
Age 25	0.1427*	0.0758	2.93%	0.0822	0.1439	2.19%
Ages 26 to 30	0.0764	0.0665	1.57%	0.0286	0.1251	0.76%
Ages 31 to 35	0.0264	0.0730	0.54%	-0.1324	0.1364	-3.52%
Ages 36 to 45	0.0838	0.0720	1.72%	-0.1190	0.1365	-3.17%
High School Information						
Known NCEA Score	0.3703***	0.0344	7.60%	0.1991***	0.0741	5.30%
Actual NCEA Score	-0.0047***	0.0002	-0.10%	-0.0024***	0.0005	-0.06%
Literacy/Numeracy	0.0819***	0.0259	1.68%	0.1404***	0.0473	3.74%
School Decile 1	0.4710***	0.0423	9.67%	0.084	0.0938	2.23%
School Decile 2	0.2370***	0.0419	4.87%	-0.1128	0.0899	-3.00%
School Decile 3	0.1823***	0.0374	3.74%	0.0225	0.0804	0.60%
School Decile 4	0.1674***	0.0326	3.44%	-0.1706**	0.0696	-4.54%
School Decile 5	0.0865**	0.0391	1.78%	-0.0602	0.0812	-1.60%
School Decile 6	0.0740**	0.0374	1.52%	-0.1361*	0.0778	-3.62%
School Decile 7	0.0886***	0.0340	1.82%	-0.1345*	0.0718	-3.58%
School Decile 8	0.1264***	0.0349	2.60%	0.0019	0.0724	0.05%
School Decile 9	0.1415***	0.0323	2.91%	-0.063	0.0664	-1.68%
School Decile 10	0.1607***	0.0289	3.30%	-0.1293**	0.0601	-3.44%
Entrance Type						
NCEA Admission	0.1752***	0.0363	3.60%	-0.0128	0.0768	-0.34%
Special Admission	0.0752***	0.0270	1.54%	-0.0827	0.0559	-2.20%
Internal	-0.0312	0.0310	-0.64%	-0.2004***	0.0653	-5.33%
External	-0.1542***	0.0274	-3.17%	-0.1752***	0.0556	-4.66%
Cambridge/IB	-0.3138***	0.0703	-6.44%	-0.3563**	0.1517	-9.48%
Course Information						
Study Hours	0.0030	0.0265	0.06%	-	-	-
Known Contact	-0.0702**	0.0346	-1.44%	-	-	-
Contact Hours	-0.0101	0.0480	-0.21%	-	-	-
Class Size/10	0.0071**	0.0028	0.10%	-	-	-
Course Size/10	-0.0015***	0.0003	-0.03%	-	-	-
Internet Content	0.0187	0.0187	0.38%	-	-	-
Level 4	-0.2765***	0.1049	-5.68%	-	-	-
Level 6	-0.0218	0.0228	-0.45%	-	-	-
Level 7	-0.1097	0.1164	-2.25%	-	-	-
Individual Academic Information						
Number of Courses	-	-	-	-0.0010	0.0119	-0.03%
6+ Courses	-	-	-	-0.7818***	0.0963	-20.81%
Double Degree	-0.1524	0.0936	-3.13%	-0.4927**	0.2423	-13.11%

One Campus	-0.0075	0.0247	-0.15%	0.0446	0.0559	1.19%
First-Year Programmes of Entry						
BA	0.0837***	0.0316	1.72%	0.3280***	0.0748	8.73%
BBus	-0.1791***	0.0469	-3.68%	-0.1242**	0.0719	-3.30%
BCIS	0.0167	0.0372	0.34%	-0.0737	0.0868	-1.96%
BCS	-0.2971***	0.0410	-6.10%	-0.1786**	0.0887	-4.75%
BDe	-0.3306***	0.0397	-6.79%	-0.1221	0.0818	-3.25%
BEdu	-0.5306***	0.0456	-10.89%	-0.2231**	0.0948	-5.94%
BEngT	0.0896*	0.0462	1.84%	-0.1237	0.1064	-3.29%
BHS	-0.3598***	0.0328	-7.39%	-0.033	0.0642	-0.88%
BIHM	-0.2196***	0.0407	-4.51%	-0.0913	0.0930	-2.43%
BMS	0.1758**	0.0727	3.61%	0.1497	0.1930	3.98%
BSR	-0.1145***	0.0366	-2.35%	0.1480*	0.0795	3.94%
Pseudo R²		0.1339			0.1063	
Log-Likelihood		-18,985.6			-4,417.61	
Area Under the ROC Curve		0.7553			0.7125	
n		50,932			9,301	

Notes: *** Indicates significance at the 1% level
** Indicates significance at the 5% level
* Indicates significance at the 10% level

Figure 1
Area Under the ROC Curve
**Course Non-Completion in the First Year**

Area under ROC curve = 0.7553
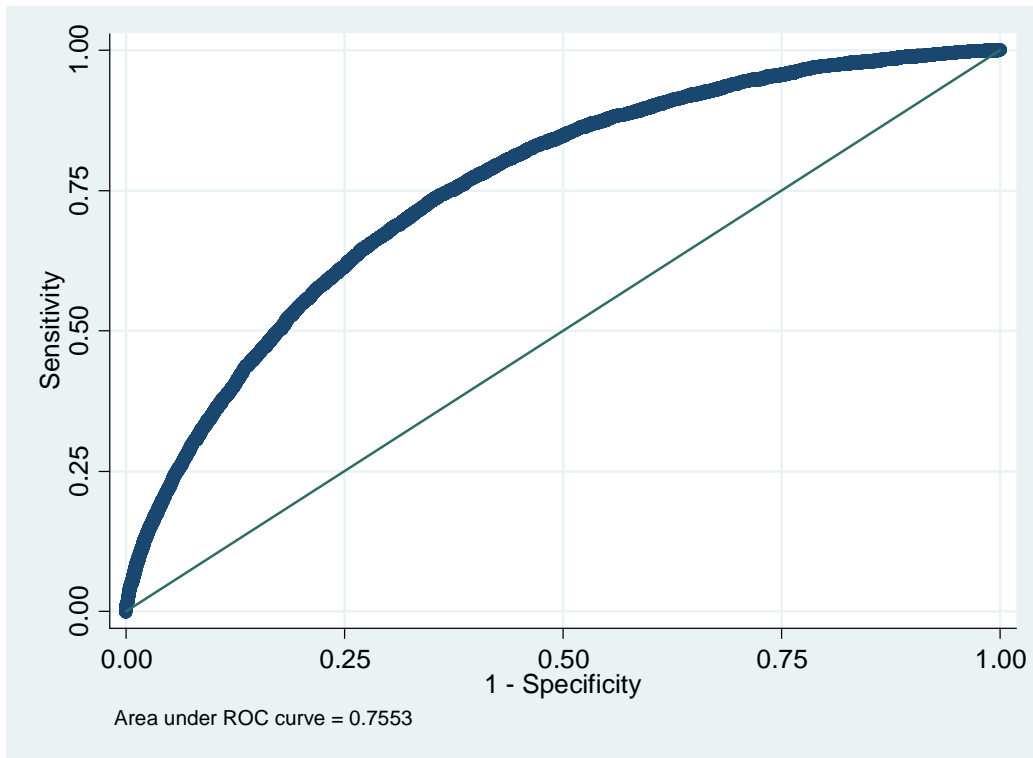
Figure 2
Area Under the ROC Curve
Student Non-Retention in the Second Year

Area under ROC curve = 0.7125
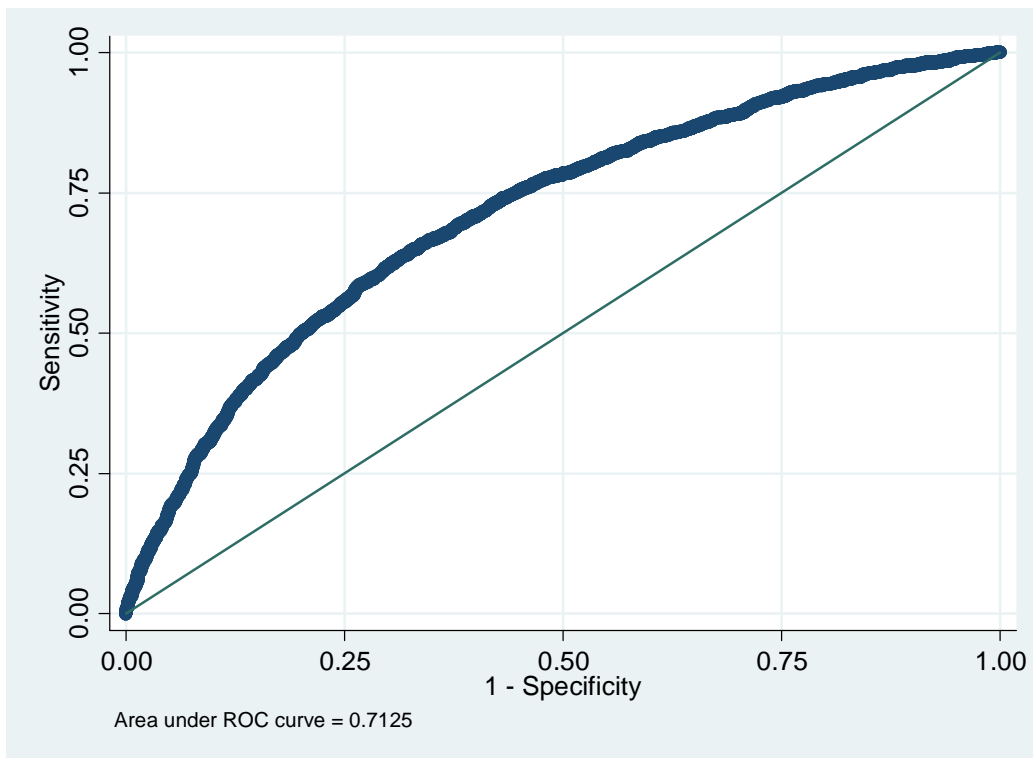
Table 3 Percentage of Outcomes Correctly Identified Validation Subsample		
	Course Non-Completion in First Year	**Student Non-Retention in Second Year**
Top 1 Decile (top 10%)	29.25%	23.33%
Top 2 Deciles (top 20%)	47.57%	40.91%
Area Under the ROC Curve	0.7553	0.7125
n	50,932	9,301