Is It Only High School Grades That Matter? Understanding Rentention In A Social Science Undergraduate Program.

By Di Wang

(Student Number: 8739458)

Major Paper presented to the
Department of Economics of the University of Ottawa
in partial fulfillment of the requirements for the M.A. Degree
Supervisor: Professor Victoria Barham

ECO 6999

Ottawa, Ontario
December 2017

Acknowledgements

This Major Research Paper has been completed with the assistance of lots of people. I really appreciate my supervisor, Professor Victoria Barham, who provided me with the research idea, as well as patient guidance and encouragement. I am also grateful to Professor Kathleen Day who provided me with important suggestions when I encountered difficulty with estimation procedure. I would like to also thank Jean-Luc Daoust for preparing that data and his support on the ethics application to access to the data. Professor Hélène Pellerin, Director of the EIL program, also provided me with information about the program. Finally, I also wish to thank Philippe Busser who helped me to get the access to the computer used for analyzing the data and facilitated access to my supervisor.

Abstract

I study factors affecting student retention in the International Studies and Modern Languages (EIL) program at University of Ottawa. Students admitted to the program have excellent high school admission averages but low rates of retention. I use uOttawa administrative data to estimate a probit model of first-year and second-year retention rates and find that the key explanatory variables differ. An unusual finding is that the high school admission average has a negative effect on both first-year and second-year retention rates of EIL students. I find that in addition to the admission average and first-semester grades, the loss of scholarship matters for 1st year retention (particularly for students from outside the National Capital Region), but not for second-year retention. Program language matters for second-year retention, but does not affect retention after the first year of study.

1. Introduction

In this paper, I investigate factors affecting student retention in an undergraduate program (International Studies and Modern Languages) at the University of Ottawa. Retention rates are calculated as the percentage of a program's full-time, first-year undergraduate students who persist in their original program of enrollment after the first academic year. Retention rates are viewed as a measure of the overall level of students' happiness and satisfaction with their academic program. However, retention rates also reflect the appropriateness of admission decisions by the university.

Academic success is typically found to be a strong predictor of student retention (Hu & John (2001); Leppel (2002); Wetzel, O'Toole & Peterson (1999)), and high school graduation averages are a strong predictor of academic success (Beattie, Laliberté & Oreopoulos (2016)). It is therefore surprising that although the admission averages for students in the EIL program are higher than for the Faculty of Social Science as a whole, the retention rate is lower. In the years from 2006 to 2012, the average retention rate in second year in the EIL program was only 66%, whereas for the Faculty of Social Science and the University of Ottawa as a whole, the rates were 76.8% and 81.8% respectively. Surprisingly, the problems with retention in EIL continue beyond second year. By the third year, only 48.6% of students who initially registered in EIL continue to pursue their studies in this program. More than half of the students have dropped or transferred out of the program.[1]

[1] Interview with Prof. Hélène Pellerin on September 7, 2017.

Given that the retention figures for the EIL program seem to be a counter-example to results reported elsewhere in the literature, there is a need for additional research to get a better understanding of why the students in the EIL program have a lower retention rate than do others at uOttawa. In this paper, I estimate a probit model to explore the specific factors that predict student persistence in the EIL program. These findings may help to (i) possibly refine admission criteria, (ii) develop targeted strategies for supporting student success, and (iii) possibly review certain aspects of the program design.

This paper proceeds as follows. In Section 2, I briefly explain the structure of the EIL program. I subsequently, in Section 3, review the academic literature on retention to identify potential explanatory variables. Section 4 discusses the data set as well as three sub-samples. Section 5 presents the methodology used for estimating retention rates. The results of the estimations are discussed in Section 6. Section 7 presents my conclusions.

2. A Brief Introduction to The EIL Program

The International Studies and Modern Languages program, known as the EIL program, is a multi-disciplinary program offered jointly by the Faculty of Social Science and the Faculty of Arts at the University of Ottawa. This program is designed to provide students with a solid and rigorous grounding in Political Science, while at the same time exposing them to a wide range of viewpoints, including economics, geography, history and world religions. Equally importantly, this program requires that students acquire a high level of competency in a second and third language. Students in the EIL program are required to be proficient in English and French and at least one additional modern language. They can choose their third

language from the following eight modern languages: Arabic, Chinese, German, Italian, Japanese, Portuguese, Russian or Spanish. Except for the required courses, the rest of compulsory and selective courses are very similar to those of Political Science students. However, unlike the standard Political Science Program, EIL student cannot choose the co-op option. In the years from 2006 to 2012, the average retention rate in second year in the Political Science program, a close substitute for EIL, was 74.73%, which is much higher than that of EIL program at 66%.

3. Literature review

There is an extensive academic literature examining factors affecting student retention in postsecondary education. Gender is generally found to be a determinant of persistence. Female students are more likely to continue in their studies (Finnie *et al* (2017); Angrist, Lang & Oreopoulos (2009); Hu & John (2001); Leppel (2002); Childs, Finnie & Martinello (2017); Bean (1980)). Researchers have argued that female students have better study habits, and are therefore more likely to do well in school than males (Angrist, Lang & Oreopoulos (2009); Beattie, Laliberté & Oreopoulos (2016)). In contrast, Herzog (2005) argues that males are more likely to transfer programs than are females, but there is no actual difference in dropout rates. Lindo, Sanders & Oreopoulos (2010) find that placing students on academic probation increases the drop-out rate for males, but does not affect females. Singell (2004) does not find a statistical difference.

First-semester academic performance is strongly correlated with retention (Murtaugh, Burns & Schuster (1999); Hu & John (2001); Leppel (2002); Lindo, Sanders & Oreopoulos

(2010); Wetzel, O'Toole & Peterson (1999); DesJardins, Ahlburg & McCall (1999); Singell (2004); Childs, Finnie & Martinello (2017); Bean (1980)). Leppel (2002) finds that academic performance reflects students' intellectual capability and a brighter student can achieve more than a less intelligent student in the same amount of study time. Beattie, Laliberté & Oreopoulos (2016) find that it is past high school grades, rather than non-academic measures, which predicts whether students will become divers (students whose first-year college average is far below expectations) or thrivers (students who exceed expectations); divers are more likely to leave than are thrivers. However, Beattie *et al* (2017) report that poor time management and less study hours are the key factors underlying poor performance, and are strongly correlated with the decision to drop out.

Many researchers have found that high school grades are a key determinant of retention in postsecondary education (Finnie *et al* (2017); Murtaugh, Burns & Schuster (1999); Lindo, Sanders & Oreopoulos (2010); DesJardins, Ahlburg & McCall (1999); Childs, Finnie & Martinello (2017); Porter & Swing (2006); Herzog (2005)). High school grades are also strongly correlated with first-semester university GPA (Beattie, Laliberté & Oreopoulos (2016)), which is regarded as the strongest factor affecting student retention.

Age also plays a significant role in determining persistence. However, there is a lack of consensus on the effect of age. Some researchers find that older students are less likely to leave without completing their degree than are younger students (Finnie *et al* (2017)). In contrast, Murtaugh, Burns & Schuster (1999), Hu & John (2001), DesJardins, Ahlburg & McCall (1999), Singell (2004), Leppel (2002) and Childs, Finnie & Martinello (2017) find

that student retention rates decrease with age. Since older students are more likely to continue employment while attending university, they have relatively less time available to spend on schoolwork than do younger students (Leppel (2002)).

There is some evidence that postsecondary students who have left home to study have lower retention rates after the first semester than do local students. Finnie *et al* (2017) and Childs, Finnie & Martinello (2017) find that the effect of not being local on student persistence is not statistically significant. In contrast, Murtaugh, Burns & Schuster (1999), Herzog (2005), Singell (2004) and DesJardins, Ahlburg & McCall (1999) all find that out-of-state (non-local) students are more likely to leave than in-state students. One factor explaining the difference in persistence between local and out-of-state students in the USA is that local students have preferential tuition fees (Herzog (2005); DesJardins, Ahlburg & McCall (1999)).

Post-secondary education is costly. A student is more likely to attend and to persist in pursuing post-secondary education if the student receives financial support from their university, family or other institutions (Bettinger *et al* (2012); Hu & John (2001); Singell (2004)). Access to financial aid also increases application rates (Bettinger *et al* (2012)). Herzog (2005) finds a positive relationship between financial aid and retention rates, but this effect only lasts for the first semester. Although there appears to be evidence that an increase in tuition will reduce retention, academic success is a more significant determinant of persistence (Wetzel, O'Toole & Peterson (1999)). Nevertheless, DesJardins, Ahlburg & McCall (1999) find that financial aid has an ambiguous influence on student retention.

Race is also related to student retention. Murtaugh, Burns & Schuster (1999) find that black, Hispanic and American Indian students are more likely to drop out than are white students in the univariate analysis, while in the multivariate analysis, black students are less likely to withdraw than are white students of the same age, GPA and residency. The gap in persistence across ethnic groups is explained by differences in academic performance (Hu & John (2001); Leppel (2002); Wetzel, O'Toole & Peterson (1999)). Asian students are less likely to drop out after one year of study than are white students (DesJardins, Ahlburg & McCall (1999)). In contrast, Singell (2004) find that race is not important in predicting student persistence.

Family income also affects student attrition. Students from high-income families are less likely to drop out than are students from middle-income and low-income families (Herzog (2005); Hu & John (2001); Leppel (2002); Singell (2004)). However, Childs, Finnie & Martinello (2017) find that family income does not have a statistically significant effect on student retention.

A number of other factors have been identified as influencing retention. Childs, Finnie & Martinello (2017) find that the students whose parents did not complete high school have a lower dropout rate since they have stronger self-selection in the decision about whether to attend university, however, aboriginal students are more likely to drop out after controlling of high school and postsecondary performance. Students who have transferred credits from another university are also more likely to persist since they have university-level experience and are more likely to be accustomed to college life (DesJardins, Ahlburg & McCall (1999)).

Living on campus increases the likelihood of retention since it promotes social integration (Herzog (2005)). Getting married and having children has a negative impact on student retention rates (Leppel (2002)). Single students are less likely to change program (Wetzel, O'Toole & Peterson (1999)). Childs, Finnie & Martinello (2017) find that students who are from single-parent families are more likely to drop out, but less likely to transfer programs.

4. Data

The data used to examine the factors affecting retention in the EIL program at the University of Ottawa are administrative data extracted by the Office of the Registrar and includes all students who registered in EIL in the fall term between 2010 and 2015. The data set includes student-level information concerning a wide variety of potential explanatory variables and follows the students for two years. There is a total of 496 students. The data set can be divided into five cohorts, reflecting the year in which the students were admitted. Of the 496 students admitted to EIL in the 6 year period, 77 observations are students admitted in the 2010 cohort, 94 observations for 2011, 90 observations for 2012, 67 observations for 2013, 73 observations for 2014, and 95 observations for 2015.

The literature review suggests a number of factors which should be considered when studying the determinants of student retention in the EIL program. Candidate factors are gender, university GPA, high school grades, age, whether the student is local or not, whether the student has earned credits before being admitted, and housing status. An additional factor that may be important in explaining retention in EIL is whether the student was born outside Canada. If so, then this individual is more likely to speak more than two languages, which

might have a positive effect on academic success in the EIL program. I also know whether a student is the first member of their family to attend university. While race was found to be an important predictor of retention in American research, it is excluded in this paper. Unfortunately, I have only limited information regarding students' family background. I do not know family income, or whether they are from a single-parent family, etc. However, I do know whether they were awarded an admission scholarship. Additionally, the language of the program in which the student is enrolled is also a potential predictor of retention, since high language competency is necessary for success in the EIL program. The category of the educational institution before entering the university is also potentially a factor which should be considered.

Table 1 reports descriptive statistics for the full sample. Each student is observed for up to two years; students who withdraw from EIL after one year are not observed in the data set for the second year. The first-year retention rate for the EIL program is 83%, which is lower than that for the university as a whole, and was 87.3% in 2013. The second-year retention rate for EIL students is 61.5%; 80.8% students were observed for two years.

80% of the students admitted to EIL are female, which is a much higher proportion than for the university as a whole (61%). The age at admission ranges from 16 years old to 40 years old; the average age is 18.90 years old. The high school admission average, the only factor used for determining eligibility for admission, ranges from 63% to 99%. No admission average was reported for four students. The mean admission average was 84%, which is higher than for the Faculty of Social Science, and is the same as for the university as a whole.

The first semester GPA variable is calculated as the average of the alphanumeric grades earned during the first semester of university studies. This variable ranged from 0 (grades of F or incomplete for all courses) to 10 (grades of A+ in all courses). The average first-semester GPA was 6.60, which is a B. No first semester GPA is reported for 27 students.[2]

Only 430 students of the 496 admitted to the program answered the question in OUAC regarding university attendance by parents, and of these only 50 students indicated that they are first generation students. Seventy percent of the students are from outside the National Capital Region. 16% of the students enter the EIL program with university-level credits earned at other universities or colleges. Approximately 45 percent of the EIL students were admitted from an Ontario high school. Fifty percent of the students lived in campus housing during the first year, whereas the rest lived off campus. As compared to the overall university population, a high proportion of EIL students lived on campus. However, given that more than 50 percent of students were from outside the National Capital Region, this means that a substantial number of non-local students lived off campus.

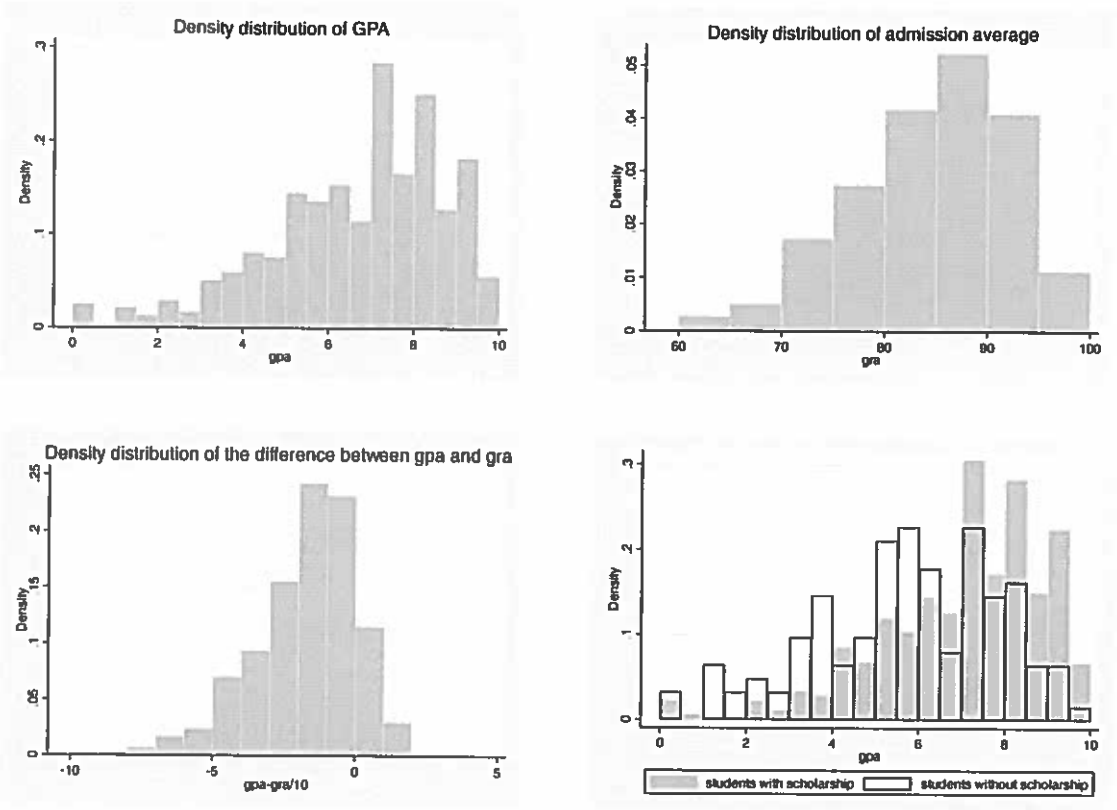
The EIL program was the first-ranked choice of 74 percent of those admitted.[3] In contrast, for the university as a whole, only 57% of the students registered in their first choice program. 32 percent of EIL students study in English, 33 percent in French, and 35 percent in

[2] 16 of these students persisted after the first year, whereas 11 students dropped out.

[3] The range of the ranking of the choice is from 1 to 5 and, surprisingly, one student who enrolled in EIL had ranked this program as their 7th choice.

French Immersion. For the university as a whole, 69% students study in English, while the rest study in French.[4]

EIL students were born in 32 countries; however, 90 percent were born in Canada, and only 6 percent are international students. The percentage of international students in the EIL program is exactly the same as for the university as a whole in 2014. "INT" (international students) and "country of birth" are highly correlated, so only country of birth is used in estimating my model.



The density distribution diagrams above show the most EIL students have a GPA between 7 and 8.5, and there is a large jump at 7; while most students have admission

[4] The university does not publish separately the proportion in French immersion.

averages between 85% and 90%, there is a sharp decline after 95%. Almost all students have a lower GPA at the end of their first semester of university study than their high school admission average. This has important financial implications. Newly-admitted students are awarded an entry scholarship if their admission average is higher than 80%; and these students retain their scholarship only if they maintain a GPA of at least 8.5 during their university studies. Thus, Figure 4 depicts the density distribution on GPA for both students who have admission scholarships and for those students who don't have admission scholarships. A higher proportion of students who have scholarships have a higher GPA; most EIL students with admission scholarships have a GPA between 7 and 7.5. The proportion of students with a high GPA among students without admission scholarships is lower than that among those with admission scholarships.

4.1 Descriptive Statistics: First-year Drop Out Subsample, Second-year Drop Out Subsample and Persistent Student Subsample

There are 83 students who dropped out after the first year and an additional 39 students who persisted after the first year but consequently dropped after the second year; only 296 students continued in the EIL program after the first two years. The means of the variables "GENDER", "country" and "INT" are similar for all three groups.

Students who persist are younger, with a higher GPA in their first session of university study and a higher admission average than those who drop out. No admission average is reported for 2 students who dropped after first year and for 2 students who persisted. No first-session GPA is reported for 11 students who dropped after their first year, and for 3 students who dropped after second year; there is no first-session GPA reported for 12

students who persist. The proportion of students who report that their parents attended university is 83.1%, 76.9% and 87.8% respectively for each sub-population. The percent of students whose parents have university degrees and who are from outside of the National Capital Region is higher for students who stayed in the EIL program than for those who dropped. The percentage of students entering the EIL program with existing university credits is highest for students who drop out after the first year. The percentage of students who are from Ontario high schools is the highest for students who drop after the second year. The percentage of students who live on campus in the first year and ranked EIL as their first choice, as well as those whose program language is French Immersion is higher for students who persisted.

This examination of the data suggests that it is appropriate to include the following variables in the estimation: the high school admission average, the university first semester grade point average, age, whether the student is the first member of their family to attend university, whether the student is from National Capital Region, whether the student comes with university credits, whether or not the student came from an Ontario high school, whether the student lives on campus, whether or not the student ranked EIL as their first-choice program, and whether the program language is Immersion.

5. Methodology

The relationship we are interested in estimating can be described as:

$$y_i = \alpha + X_i\beta_i + \varepsilon_i, \ for\ i = 1, 2, \ldots., N. \qquad (1)$$

where y_i is a binary dependent variable, and is equal to 1 if the individual remains enrolled in the EIL program, 0 if the individual drops out or switches to other programs; X_i represents the set of explanatory variables, which includes all possible predictors of student retention; ε_i is the error term.

OLS estimation is not well-suited to estimating an empirical model when the dependent variable is bounded between zero and one. The estimated values $\widehat{y_i}$ values obtained from OLS estimation will often be out-of-range, that is, the estimated probability will not be in the interval [0,1]. This problem arises because of linear extrapolation errors, which results in inaccurate predictions of probabilities (Pohlman, J. T. & Leitner (2003)). Instead of using OLS, I use the probit model, which ensures that the predicted probability of persistence stays between [0,1].

For the probit model, the probability that $y_i = 1$ is the same as the expected value of y_i:

$$P(y_i = 1|X_i, \varepsilon_i) = E(y_i|X_i, \varepsilon_i) = G[(X_i\beta_i + \varepsilon_i)] \qquad (2)$$

where G is a standard normal cumulative distribution function (CDF) with strict values range between 0 and 1 for all real number r.

Since the model estimated in this paper is a non-linear regression model, the coefficients of the regressors do not have a direct interpretation because the scale factor depends on all of the explanatory variables.[5] Instead, to interpret the effect of individual covariates it is

[5] Jeffrey M. Wooldridge. *Introductory Econometrics: A Modern Approach (5th Edition).* 589-593.

necessary to calculate the marginal effects for each variable. However, the procedure for calculating marginal effects is different for continuous and for dichotomous variables.

The partial effect of continuous variables on student retention probability can be derived from the partial derivative. The marginal effect of a continuous variable x_i on the probability $P(y_i = 1|X_i, \varepsilon_i)$, holding all other variables constant, can be derived as:

$$\frac{\partial p(X_i)}{\partial x_i} = g(X_i\beta_i + \varepsilon_i)\beta_i, \qquad where\ g(r) \equiv \frac{dG}{dr}(r)$$

$G(\cdot)$ is strictly increasing CDF, and $g(r) > 0$ for all r. The partial effect always has the same sign as β_i.

The partial effect of a dichotomous variable is estimated differently from that of a continuous variable. The partial effect of changing x_i, a dichotomous variable, from zero to one, holding all other variables fixed, is derived as:

$$\Delta = G(\bar{x}\beta_i, d = 1) - G(\bar{x}\beta_i, d = 0)$$

6. Results

6.1 First Year Retention

I first estimate equation (2) with the first year retention rate as the dependent variable. Table 3 shows the estimated average marginal effects of each covariate for the first year retention rate for a number of different specifications. In each of the specifications, the high school admission average (ADM_AVG), the first semester grade point average (GPA) and whether or not the student came from an Ontario high school (CATEGORY) are statistically significant. The estimated coefficients are similar for the different specifications although the

coefficient estimate of CATEGORY is much larger than that of GPA for specifications (1) to (4).

The baseline specification is similar to the standard approach in the literature. In the baseline specification, I predict the first-year retention rate using GENDER, GPA, ADM_AVG, CATEGORY, AGE, whether the student is from National Capital Region (REGION), whether the student is the first member attending university (GEN1), whether the student comes with university credits (CREDITS), whether the student lives on campus (REZ), whether the student was born in Canada (COUNTRY), whether or not the student ranked EIL as their first-choice program (CHOICE), whether the program language is English (LAN_ENG), whether the program language is French (LAN_FRE), whether the student is admitted in 2010 (COHORT2010), whether the student is admitted in 2011 (COHORT 2011), whether the student is admitted in 2012 (COHORT 2012), whether the student is admitted in 2013 (COHORT2013), and whether the student is admitted in 2014 (COHORT 2014) as control variables.

The fact that GPA, ADM_AVG, AGE and CATEGORY are statistically significant is consistent with the findings of Murtaugh, Burns & Schuster (1999), DesJardins, Ahlburg & McCall (1999), Lindo, Sanders & Oreopoulos (2010), Beattie, Laliberté & Oreopoulos (2016), Childs, Finnie & Martinello (2017) and Finnie *et al* (2017). Although the coefficient estimate of CHOICE is statistically significant, its average marginal effect is not significant. The sign of the coefficient estimate of GPA is positive, which implies that a higher university GPA increases the probability of retention. However, ADM_AVG, CATEGORY and AGE

have negative effects on retention. A higher admission average reduces the probability of retention, which is an unusual result and in sharp contrast to what is found elsewhere in the literature.[6] Students who are not from Ontario high schools are more likely to drop out. Older students are less likely to persist, which is consistent with the finding of Murtaugh, Burns & Schuster (1999), Hu & John (2001), DesJardins, Ahlburg & McCall (1999), Singell (2004), Leppel (2002) and Childs, Finnie & Martinello (2017).

I hypothesize that the loss of the admission scholarship may be an important influence on first-year retention. Although I do not observe financial information for each student, an admission scholarship is awarded to students who have an admission average of 80, and this financial support continues if the student GPA is greater than 8.5 every year. I therefore construct a dichotomous variable "lose scholarship" (LOSESCH) which is equal to 1 if the student's GPA≥8.5 and ADM_AVG-GPA<0, but is 0 otherwise. I introduce LOSESCH in specification (3). It is interesting to see that although LOSESCH is not statistically significant, the estimated coefficient is larger than that for ADM_AVG, which suggests it may be an important factor in retention.

One fact that the estimated coefficient is not statistically significant may be due to the fact that data set is not large enough. An additional fact which should be taken into consideration

[6] If we control university GPA and high school admission averages in the probit model for retention estimation at the same time, the difference between GPA and ADM_AVG is also a potential indicator for retention automatically. The coefficient estimates of this difference and ADM_AVG jointly affect the coefficient estimate of real ADM_AVG. If the former is positive and has a larger effect than the latter, then the coefficient estimate of ADM_AVG could be negative. In this paper, the negative estimated coefficient of admission averages can be explained by this case. $y_i = \alpha_i + \beta_1(GPA - ADM_{AVG}) + \beta_2 GPA + \beta_3 ADM_{AVG} + \varepsilon_i = \alpha_i + (\beta_1 + \beta_2)GPA + (\beta_3 - \beta_1)ADM_{AVG} + \varepsilon_i$

is that students from outside the National Capital Region face higher costs in pursuing their education at uOttawa than do local students who can live at home. For these students, the loss of a scholarship may be particularly burdensome. Consequently, I interact REGION&LOSE SCHOLARSHIP in specification (4) which is my preferred specification. In this specification, the coefficient estimates of GPA, ADM_AVG, REGION, CATEGORY, LOSESCH and the REG_LOSESCH interaction term are all statistically significant, while the average marginal effect of GPA, ADM_AVG and CATEGORY are statistically significant.

On average, CATEGORY has the biggest marginal effect on the probability of first-year retention with a coefficient estimate of -0.0944, followed by CHOICE with a coefficient estimate of 0.0688.

On average, an increase of 1 unit in first-semester GPA increases the probability of retention for this individual by 5.01 percentage points; however, and surprisingly, an increase of 1 unit in high school admission average decreases the probability of persistence by 0.71 percentage points; if the individual changes from an Ontario high school to a non-Ontario high school, the probability of falls by 9.44 percentage points.

I undertake a number of diagnostic tests to ensure that the reported estimation results are valid. I implement the VIF test to check for multicollinearity, and find that there is no problem of multicollinearity. The Wald test for joint significance shows that we can reject the null hypothesis that the coefficient estimates of all 19 variables are zero. Furthermore, the result on comparing fitted probability and dichotomous outcome shows that the percentage of correctly specified values is 86.62%. In this model, there are 53 observations misclassified as

persistence when the correct classification is dropping, and 2 values misclassified as dropping when the correct value is retention. The remaining 356 observations are correctly predicted.

6.2 Second Year Retention

Whereas the existing literature focuses on first-year retention, in this paper I also consider second-year retention. Students whose first-year GPA falls below 5.0 (a C+) are placed on academic probation, and must improve their performance in second year. The estimated average marginal effects of each covariate for the second year retention rate are presented in Table 4.

The baseline specification is the same as the baseline specification for first-year retention. Notice that n=329 rather than 411 as I do not have data on second year retention for students who began in 2015. It can be seen that GPA, ADM_AVG, AGE and COUNTRY are statistically significant, which is also confirmed by Hu & John (2001), Lindo, Leppel (2002), Sanders & Oreopoulos (2010), Beattie, Laliberté & Oreopoulos (2016), Childs, Finnie & Martinello (2017) and Finnie *et al* (2017). The sign of GPA is positive, while the sign of ADM_AVG, AGE and COUNTRY are negative. Interestingly, ADM_AVG has a negative effect on both first and second year retention.[7]

After dropping AGE, COUNTRY is not significant although both GPA and ADM_AVG are still statistically significant in specification (2).

[7] The reason why high school admission averages have a negative effect on retention is the same as I explain in 6.1.

I try to explore the effect of language on retention since the EIL program requires higher linguistic competency than does other programs. Students whose program language is French Immersion have higher first-semester GPA in university study and a higher probability of retention than those whose program language is French or English. Thus, I add two GPA& LANGUAGE interactions in specification (3). I find that GPA, ADM_AVG, LAN_ENG and the GPA_LAN_ENG interaction terms are statistically significant. But the VIF test on multicollinearity reports that LAN_ENG, LAN_FRE and the two interaction terms are highly correlated.

In order to address the problem of multicollinearity, I introduce GPA SQUARE (GPAS) and two interaction terms of GPAS&LANGUAGE instead of GPA&LANGUAGE in specification (4). I find that LAN_ENG becomes insignificant, while GPA, ADM_AVG and GPAS_LAN_ENG are still statistically significant. Nevertheless, the VIF test on multicollinearity shows that GPA and GPAS are highly correlated. The correlation coefficient between GPA and GPAS is 0.9725, therefore, I drop GPAS in specification (5) which is my preferred specification. In this specification, the coefficient estimates of GPA, ADM_AVG, LAN_ENG and the GPAS_LAN_ENG interaction term are statistically significant, while only the average marginal effect of GPA, ADM_AVG and COUNTRY are statistically significant.

On average, GPA and COUNTRY have the largest average marginal effect on second-year retention with coefficient estimate of 0.1344 and -0.1056 respectively. However, there are very few significant variables both in probit estimation and the estimation of the

average marginal effect; to obtain a more precise estimation would require a larger sample size.

On average, if the student's GPA increases by 1 unit, the probability of retention for this individual will increase by 13.44 percentage points; however, if the student's admission average increases by 1 unit, the probability on persistence for this individual will decrease by 0.79 percentage points; if the student changes his or her born country from Canada to non-Canada, the probability of retention decreases by 10.56 percentage points.

In order to check the validity of the final model, I implemented a number of diagnostic tests. The VIF test of multicollinearity shows that there is no multicollinearity problem in this model. The null hypothesis of the Wald test on the joint significance of all eighteen explanatory variables is rejected: the p-value of Wald test is 0.0002, which means that we can reject the null hypothesis that they are all zero. Moreover, if I compare the fitted probability and dichotomous outcome, I find that the percentage of correctly specified values is 81.16%. In this model, there are 53 observations misclassified as persistence when the correct classification is dropping, and 9 values misclassified as dropping when the correct value is retention. The remaining 267 observations are correctly predicted.

6.3 Robustness Tests

As an additional check on my finding that the factors determining first and second-year retention rates are different, I estimate RET1 using my preferred specification for RET2, and I estimate RET2 using me preferred specification for RET1. These results are reported in

Table 6. Column (1) and (3) reproduce the estimated coefficient for RET1 and RET2 using the preferred specifications previously reported in Table 3 and Table 4 respectively.

The results reinforce my claim that although the first-semester GPA and the admission average always matter for retention, that whether or not the student is from outside the National Capital Region, whether the student is from an Ontario high school and whether the student loses their scholarship matters only for the first year retention rate. In contrast, whether the program language is English influences persistence beyond second year but does not affect first-year retention rate.

7. Conclusions

This paper examines the factors affecting student retention in the EIL program at uOttawa. Given that EIL students have a higher admission average than uOttawa students in general, but a lower retention rate, it is interesting to investigate the factors affecting student retention in this program. Using uOttawa administrative data on students' academic and personal information, I estimate a probit model of first-year and second-year retention rates.

Surprisingly, I find that the admission average has a negative effect on both first and second year retention rate; this is unlike other findings in the literature. I find that the loss of scholarship negatively influences the decision to persist in the EIL program beyond first-year retention, especially for students who are from outside the National Capital Region. However, this does not affect second-year retention. Program language is crucial for second-year retention, but does not affect retention after the first year of study.

These findings point to some possible strategies for improving the retention rate in the EIL program. The above results underscore the key role of financial support, especially for those who are from outside the National Capital Region. It would also be important for the EIL Program Director to examine why program language matters: one possibility is there something different about the experience of students studying in English rather than in French or French Immersion that could explain the problem with retention for these students.

For future research regarding retention in EIL this program, it would be interesting to estimate the third-year student retention rate, since the third-year retention rate may be highly correlated with the graduation rate. Also, endogeneity issues in my model should be investigated. For instance, the university first semester GPA has a positive effect on student retention, however, a student who is dissatisfied with their choice of program may study less, which in turn lowers the university GPA. Thus, more research should be done to explore this issue.

References

Angrist, J., Lang, D., & Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics, 1* (1), 136-163.

Beattie, G., Laliberté, J. W. P., Michaud-Leclerc, C., & Oreopoulos, P. (2017). What Sets College Thrivers and Divers Apart? A Contrast in Study Habits, Attitudes, and Mental Health. National Bureau of Economic Research: W23588.

Beattie, G., Laliberté, J. W. P., & Oreopoulos, P. (2016). Thrivers and Divers: Using Non-Academic Measures to Predict College Success and Failure. National Bureau of Economic Research: W22629.

Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education, 12* (2), 155-187.

Bettinger, E. P., Long, B. T., Oreopoulos, P., & Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. *The Quarterly Journal of Economics, 127* (3), 1205-1242.

Childs, S. E., Finnie, R., & Martinello, F. (2017). Postsecondary Student Persistence and Pathways: Evidence from the YITS-A in Canada. *Research in Higher Education, 58* (3), 270-294.

Dekker, G., Pechenizkiy, M., & Vleeshouwers, J. (2009). Predicting students drop out: A case study. *Educational Data Mining 2009.*

DesJardins, S. L., Ahlburg, D. A., & McCall, B. P. (1999). An event history model of student departure. *Economics of Education Review, 18* (3), 375-390.

Finnie, R., Fricker, T., Bozkurt, E., Poirier, W., & Pavlic, D. (2017). Using Predictive Modelling to Inform Early Alert and Intrusive Advising Interventions and Improve Retention.

Herzog, S. (2005). Measuring determinants of student return vs. dropout/stopout vs. transfer: A first-to-second year analysis of new freshmen. *Research in Higher Education, 46* (8), 883-928.

Hu, S., & John, E. P. S. (2001). Student persistence in a public higher education system: Understanding racial and ethnic differences. *The Journal of Higher Education, 72* (3), 265-286.

Leppel, K. (2002). Similarities and differences in the college persistence of men and women. *The Review of Higher Education, 25* (4), 433-450.

Lindo, J. M., Sanders, N. J., & Oreopoulos, P. (2010). Ability, gender, and performance standards: Evidence from academic probation. *American Economic Journal: Applied Economics*, *2* (2), 95-117.

Murtaugh, P. A., Burns, L. D., & Schuster, J. (1999). Predicting the retention of university students. *Research in Higher Education*, *40* (3), 355-371.

Porter, S. R., & Swing, R. L. (2006). Understanding how first-year seminars affect persistence. *Research in Higher Education*, *47* (1), 89-109.

Pohlman, J. T., & Leitner, D. W. (2003). A comparison of ordinary least squares and logistic regression. *The Ohio Journal of Science*, *103* (5), 118-125.

Singell, L. D. (2004). Come and stay a while: does financial aid effect retention conditioned on enrollment at a large public university? *Economics of Education Review*, *23* (5), 459-471.

Styron Jr, R. (2010). Student satisfaction and persistence: Factors vital to student retention. *Research in Higher Education Journal*, *6* (1).

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

Appendix

Table 1 Data description for full sample

Variable	Definition	Mean	Std. Dev.	Sample size	All university sample
Dependent variables					
RET1	1 = individual remained in the EIL program after the first year, 0 = otherwise	0.83		496	87.3% (2013)
RET2	1 = individual remained in the EIL program after the second year, 0 = otherwise	0.76		401	
Independent variables					
GENDER	1 = female, 0 = male	0.80		496	61% (Fall2014)
AGE	Age at admission	18.90	2.11	496	
ADM_AVG	Admission average from high school, CEGEP, college	83.99	7.56	492	83.9%(overall); 82.5%(social science)
GPA	Grade point average after first session of university study	6.60	2.07	469	
GEN1	1 = individual is the first member of their family to take university-level studies, 0 = otherwise	0.12		430	
REGION	1 = individual is from outside National Capital Region, 0 = otherwise	0.70		496	
CREDITS	1 = individual enters EIL program with university credits from other universities or colleges, 0 = otherwise	0.16		496	
CATEGORY	1 = individual is not admitted from Ontario high school, 0 = otherwise	0.45		496	
REZ	1 = lives in university residence, 0 = otherwise	0.51		496	37.1%(Fall2014)
CHOICE	1 = EIL program ranked as first choice in OUAC, 0 = otherwise	0.74		496	57%(Fall2015)
LAN_ENG	1= program language is English, 0= otherwise	0.32		496	
LAN_FRE	1= program language is French, 0= otherwise	0.33		496	
LAN_IMMERSION	1= program language is Immersion, 0= otherwise	0.34		496	
COUNTRY	1= born in Canada, 0= otherwise	0.90		496	
INT	1= individual is not international student, 0= otherwise	0.94		496	94% (Fall 2014)
COHORT2010	1= admitted in 2010; 0= otherwise	0.16		496	
COHORT2011	1= admitted in 2011; 0= otherwise	0.19		496	
COHORT2012	1= admitted in 2012; 0= otherwise	0.18		496	
COHORT2013	1= admitted in 2013; 0= otherwise	0.14		496	
COHORT2014	1= admitted in 2014; 0= otherwise	0.15		496	
COHORT2015	1= admitted in 2015; 0= otherwise	0.19		496	

Notes: Last column is the percentage of all university sample when the dummy variable is equal to 1.

Source of the last column: https://www.uottawa.ca/institutional-research-planning/resources/facts-figures/cudo/2015-menu

Table 2 Descriptive Statistics: First-year Drop Out, Second-year Drop Out and Persistent Student

Variable	Drop 1	Drop 2	Persist
Dependent variables			
RET1	0	1	1
	(N=83)	(N=39)	(N=296)
RET2	0.12	0	1
	(N=66)	(N=39)	(N=296)
Independent variables			
GENDER	0.80	0.77	0.80
	(N=83)	(N=39)	(N=296)
AGE	19.18	19.03	18.92
	(N=83)	(N=39)	(N=296)
ADM_AVG	82.56	81.77	84.52
	(N=81)	(N=39)	(N=294)
GPA	5.28	5.23	7.03
	(N=72)	(N=36)	(N=284)
GEN1	0.13	0.17	0.10
	(N=69)	(N=30)	(N=260)
REGION	0.71	0.64	0.71
	(N=83)	(N=39)	(N=296)
CREDITS	0.22	0.10	0.17
	(N=83)	(N=39)	(N=296)
CATEGORY	0.61	0.36	0.46
	(N=83)	(N=39)	(N=296)
REZ	0.43	0.46	0.51
	(N=83)	(N=39)	(N=296)
CHOICE	0.67	0.64	0.77
	(N=83)	(N=39)	(N=296)
LAN_ENG	0.42	0.38	0.31
	(N=83)	(N=39)	(N=296)
LAN_FRE	0.29	0.36	0.35
	(N=83)	(N=39)	(N=296)
LAN_IMMERSION	0.29	0.26	0.34
	(N=83)	(N=39)	(N=296)
COUNTRY	0.90	0.90	0.89
	(N=83)	(N=39)	(N=296)
INT	0.95	0.97	0.94
	(N=83)	(N=39)	(N=296)
COHORT2010	0.14	0.13	0.20
	(N=83)	(N=39)	(N=296)
COHORT2011	0.19	0.28	0.23

	(N=83)	(N=39)	(N=296)
COHORT2012	0.18	0.28	0.22
	(N=83)	(N=39)	(N=296)
COHORT2013	0.16	0.18	0.16
	(N=83)	(N=39)	(N=296)
COHORT2014	0.12	0.13	0.20
	(N=83)	(N=39)	(N=296)
COHORT2015	0.20	0	0
	(N=83)	(N=39)	(N=296)

Notes: All numbers are the means of each variable. N is the number of the observations.

Table 3 First Year Retention

	(1)	(2)	(3)	(4)
Variables				
GENDER	-0.0022	0.0081	0.0093	0.0126
	(0.0421)	(0.0435)	(0.0435)	(0.0429)
GPA	0.0506***	0.0459***	0.0494***	0.0501***
	(0.0090)	(0.0089)	(0.0096)	(0.0095)
ADM_AVG	-0.0089***	-0.0066**	-0.0080**	-0.0071**
	(0.0031)	(0.0030)	(0.0033)	(0.0033)
AGE	-0.0371**			
	(0.0146)			
REGION	-0.0365	-0.0466	-0.0459	-0.0269
	(0.0431)	(0.0422)	(0.0419)	(0.0462)
GEN1	-0.0097	-0.0214	-0.0256	-0.0195
	(0.0512)	(0.0536)	(0.0542)	(0.0527)
CREDITS	0.0018	-0.0311	-0.0291	-0.0356
	(0.0574)	(0.0625)	(0.0621)	(0.0622)
REZ	0.0181	0.0381	0.0345	0.0274
	(0.0445)	(0.0443)	(0.0443)	(0.0432)
CATEGORY	-0.0750*	-0.0961**	-0.0948**	-0.0944**
	(0.0443)	(0.0443)	(0.0441)	(0.0438)
CHOICE	0.0724	0.0677	0.0658	0.0688
	(0.0448)	(0.0452)	(0.0450)	(0.0446)
COUNTRY	-0.0702	-0.0528	-0.0485	-0.0675
	(0.0493)	(0.0543)	(0.0555)	(0.0486)
LOSESCH			0.0420	0.0464
			(0.0436)	(0.0450)
LAN_FRE	0.0090	0.0172	0.0223	0.0260
	(0.0456)	(0.0457)	(0.0453)	(0.0439)
LAN_ENG	-0.0171	-0.0181	-0.0191	-0.0224
	(0.0429)	(0.0435)	(0.0434)	(0.0427)
LAN_IMMERSION				
COHORT2010	-0.0056	0.0006	0.0063	0.0065
	(0.0597)	(0.0594)	(0.0582)	(0.0575)
COHORT2011	-0.0055	-0.0116	-0.0060	0.0067
	(0.0534)	(0.0545)	(0.0538)	(0.0513)
COHORT2012	0.0538	0.0433	0.0505	0.0575
	(0.0464)	(0.0486)	(0.0477)	(0.0459)
COHORT2013	-0.0180	-0.0190	-0.0125	-0.0038
	(0.0616)	(0.0628)	(0.0616)	(0.0586)
COHORT2014	0.0164	0.0164	0.0225	0.0268
	(0.0546)	(0.0553)	(0.0542)	(0.0523)

COHORT2015				
REG_LOSESCH				
N	411	411	411	411

Notes: Column (1) is the baseline specification. Column (2) is the specification without AGE. Column (3) is the specification with LOSE SCHOLARSHIP. Column (4) is my preferred specification which includes the REGION&LOSE SCHOLARSHIP interaction term. *significant at 10% level. **significant at 5% level. ***significant at 1% level.

Table 4 Second Year Retention

	(1)	(2)	(3)	(4)	(5)
Variables					
GENDER	0.0445	0.0431	0.0438	0.0387	0.0387
	(0.0540)	(0.0554)	(0.0547)	(0.0542)	(0.0542)
GPA	0.0872***	0.0795***	0.0784***	0.1344**	0.1344**
	(0.0113)	(0.0113)	(0.0114)	(0.0523)	(0.0523)
ADM_AVG	-0.0126***	-0.0091**	-0.0087**	-0.0079*	-0.0079*
	(0.0040)	(0.0041)	(0.0041)	(0.0042)	(0.0042)
AGE	-0.0665***				
	(0.0202)				
REGION	-0.0152	-0.0443	-0.0480	-0.0527	-0.0527
	(0.0571)	(0.0549)	(0.0539)	(0.0533)	(0.0533)
GEN1	0.0272	-0.0013	0.0085	0.0103	0.0103
	(0.0606)	(0.0658)	(0.0643)	(0.0635)	(0.0635)
CREDITS	0.0298	-0.0493	-0.0441	-0.0408	-0.0408
	(0.0722)	(0.0810)	(0.0798)	(0.0794)	(0.0794)
REZ	-0.0089	0.0234	0.0235	0.0256	0.0256
	(0.0550)	(0.0548)	(0.0536)	(0.0535)	(0.0535)
CATEGORY	-0.0125	-0.0360	-0.0454	-0.0372	-0.0372
	(0.0523)	(0.0535)	(0.0537)	(0.0539)	(0.0539)
CHOICE	0.0697	0.0651	0.0753	0.0632	0.0632
	(0.0516)	(0.0531)	(0.0534)	(0.0531)	(0.0531)
COUNTRY	-0.1353**	-0.0913	-0.1072*	-0.1056*	-0.1056*
	(0.0556)	(0.0687)	(0.0649)	(0.0634)	(0.0634)
LAN_FRE	-0.0511	-0.0389	-0.0287	-0.0292	-0.0292
	(0.0584)	(0.0592)	(0.0620)	(0.0606)	(0.0606)
LAN_ENG	-0.0494	-0.0458	-0.0357	-0.0427	-0.0427
	(0.0547)	(0.0558)	(0.0578)	(0.0581)	(0.0581)
LAN_IMMERSION					
COHORT2010	-0.0614	-0.0580	-0.0715	-0.0661	-0.0661
	(0.0805)	(0.0821)	(0.0838)	(0.0834)	(0.0834)
COHORT2011	-0.0813	-0.0939	-0.1169	-0.1157	-0.1157
	(0.0727)	(0.0748)	(0.0760)	(0.0761)	(0.0761)
COHORT2012	-0.0627	-0.0843	-0.1030	-0.1003	-0.1003
	(0.0728)	(0.0758)	(0.0777)	(0.0779)	(0.0779)
COHORT2013	-0.1039	-0.1112	-0.1170	-0.1157	-0.1157
	(0.0820)	(0.0850)	(0.3005)	(0.0856)	(0.0856)

COHORT2014					
GPA_LAN_FRE					
GPA_LAN_ENG					
GPAS(gpa^2)				-0.0050	-0.0050
				(0.0045)	(0.0045)
GPAS_LAN_FRE					
GPAS_LAN_ENG					
N	329	329	329	329	329

Notes: Column (1) is the baseline specification. Column (2) is the specification without AGE. Column (3) is the specification with GPA&LANGUAGE interaction terms. Column (4) is the specification with GPA SQUARE and GPAS&LANGUAGE interaction terms. Column (5) is my preferred specification which is without GPA SQUARE. *significant at 10% level. **significant at 5% level. ***significant at 1% level.

Table 5 Comparison of Three Language Groups

	RET1	RET2	GPA
English(LAN_ENG=1)	0.7826	0.7015	6.5074
French(LAN_FRE=1)	0.8537	0.7626	6.3487
Immersion(LAN_IMMERSION =1)	0.8596	0.8125	6.9230

Notes: All above numbers are the means of each variable.

Table 6 Robustness Tests

	First Year Retention		Second Year Retention	
	(1)	(2)	(3)	(4)
Variables				
GENDER	0.0621	0.0365	0.1587	0.1812
	(0.2085)	(0.2063)	(0.2090)	(0.2093)
GPA	0.2517***	0.3193***	0.4601***	0.3505***
	(0.0504)	(0.0712)	(0.0871)	(0.0583)
ADM_AVG	-0.0357**	-0.0286*	-0.0334**	-0.0464**
	(0.0168)	(0.0149)	(0.0170)	(0.0187)
AGE				
REGION	-1.0183***	-0.2867	-0.2155	-0.3796
	(0.3619)	(0.2288)	(0.2350)	(0.3108)
GEN1	-0.0949	-0.0775	0.0440	-0.0142
	(0.2482)	(0.2432)	(0.2661)	(0.2665)
CREDITS	-0.1693	-0.1223	-0.1623	-0.1794
	(0.2801)	(0.2753)	(0.3034)	(0.3031)
REZ	0.1364	0.1836	0.1030	0.0566
	(0.2135)	(0.2095)	(0.2181)	(0.2209)
CATEGORY	-0.4557**	-0.4324**	-0.1678	-0.1118
	(0.2062)	(0.2020)	(0.2159)	(0.2141)
CHOICE	0.3218	0.3106	0.2623	0.2305
	(0.1961)	(0.1928)	(0.1987)	(0.1980)
COUNTRY	-0.4017	-0.3237	-0.5119	-0.4378
	(0.3500)	(0.3433)	(0.3766)	(0.3778)
LOSESCH	-0.7851**			0.0362
	(0.3885)			(0.3515)
LAN_FRE	0.1336	0.5989	0.4104	-0.1057
	(0.2310)	(0.4232)	(0.4622)	(0.2319)
LAN_ENG	-0.1102	0.5643	0.8159*	-0.1838
	(0.2067)	(0.4283)	(0.4638)	(0.2138)
LAN_IMMERSION				
COHORT2010	0.0329	0.0196	-0.2696	-0.2367
	(0.2953)	(0.2876)	(0.3096)	(0.3055)
COHORT2011	0.0338	-0.0711	-0.4482	-0.3645
	(0.2629)	(0.2557)	(0.2779)	(0.2728)
COHORT2012	0.3178	0.2414	-0.3958	-0.3213
	(0.2811)	(0.2749)	(0.2841)	(0.2765)
COHORT2013	-0.0189	-0.0503	-0.4391	-0.4153
	(0.2909)	(0.2874)	(0.3009)	(0.2954)
COHORT2014	0.1413	0.1165		
	(0.2892)	(0.2838)		

COHORT2015				
REG_LOSESCH	1.3071***			0.3793
	(0.4191)			(0.3861)
GPAS_LAN_FRE		-0.0116	-0.0123	
		(0.0084)	(0.0093)	
GPAS_LAN_ENG		-0.0143*	-0.0221**	
		(0.0082)	(0.0091)	
N	411	411	329	329
LR chi2(#)	49.01	40.75	59.87	56.80
Prob>chi2	0.0002	0.0026	0.0000	0.0000
Pseudo R^2	0.1420	0.1180	0.1719	0.1631
Log-likelihood	-148.0975	-152.2291	-144.1971	-145.7319
#(number of variables)	19	19	18	18

Notes: Column (1) is the preferred specification for first-year retention. Column (3) is the preferred specification for second-year retention. Column (2) is the estimation on first-year retention using preferred specification for second-year retention. Column (4) is the estimation on second-year retention using preferred specification for first-year retention. *significant at 10% level. **significant at 5% level. ***significant at 1% level.