Understanding the Determinants of Persistence and Academic Success in University: An Exploration of Data from Four Ontario Universities

Prepared by Martin D. Dooley, A. Abigail Payne, A. Leslie Robb, McMaster University
for the Higher Education Quality Council of Ontario

Higher Education Quality Council of Ontario
An agency of the Government of Ontario

Disclaimer:

Cite this publication in the following format:

Dooley, M. D, Payne, A. A., Robb, L. A., McMaster University (2011). *Understanding the Determinants of Persistence and Academic Success in University: An Exploration of Data from Four Ontario Universities.* Toronto: Higher Education Quality Council of Ontario.

Published by:

The Higher Education Quality Council of Ontario

1 Yonge Street, Suite 2402
Toronto, ON Canada
M5E 1E5
Phone: (416) 212-3893
Fax: (416) 212-3899
Web: www.heqco.ca
E-mail: info@heqco.ca

The authors would like to acknowledge the very helpful comments of Ursula McCloy and Huizi Zhao, and the excellent research assistance of Leyden Martinez-Fonte, Cristina Sechel and Linda Jonker.

Table of Contents

List of Tables

List of Figures

Executive Summary

We address a series of questions concerning academic persistence and success among university students using administrative data that have been collected on students in four Ontario universities and linked with information on the students' individual characteristics (including high school performance), neighbourhood, and high school. The students came directly from an Ontario high school and entered one of our four universities for full-time degree study in the fall of 1994 through 2006. The measures of persistence are the cumulative grade average and credits completed at the end of the second year in university, departures during the first two years, and degrees completed within six years. The following lessons can be inferred from our analyses.

First, the time trends reveal that the values of all four outcomes have generally been quite stable over time at each of our institutions. This stability over time in the levels of our measures of academic success in university is true of both the simple means of the variables and when we control statistically for a wide variety of individual, neighbourhood and school characteristics.

Second, academic performance in secondary school is strongly linked to all of our measures of university performance. These links are strong in the sense of both the magnitude and the precision of the estimated coefficients. Furthermore, the explanatory power of the high school grade average greatly dominates that of other variables such as university program, gender, neighbourhood average income, and average high school performance on Grade 9 EQAO tests. Understanding what lies behind the large estimated impact of high school grades is clearly important for understanding the determinants of university success. Many background factors undoubtedly contribute to academic success in both high school and university including health, parental education and income, and the secondary schooling context. This policy concern highlights the importance of linking our current data sets to other data, such as from the Ministry of Education, in order to shed light on the factors associated high school grades.

Third, the neighbourhood characteristics used in this study, including average income and others, have relatively weak links with our measures of persistence. In contrast, Dooley, Payne and Robb (2009) reported that students from low income neighbourhoods are 13 per cent less likely to apply to university than those in high income neighbourhoods (controlling for other factors including GPA) and that this gap in application rates over the last decade has remained relatively constant. Hence, neighbourhood socio-economic background appears to play an important role in gaining access to university but a more minor role in academic success once a student has registered. We hasten to add, however, that numerous small differences can still have a large cumulative impact on the life of a young person. Furthermore, there is much variation in income and other socio-economic characteristics within neighbourhoods. Differences in individual family income may have substantially more success in explaining university persistence than do differences in average income and education across neighbourhoods.

Fourth, our results point to the advantages of rich administrative data. Such data not only provide a very large sample size but also suffer much less than do survey data from response and selection bias. For example, one could do much additional research by examining

additional outcome measures and conducting more specific analysis by program, academic level, gender, mother tongue, and type of high school. As indicated above, linking our current data with additional data sets will expand the research opportunities even further.

Introduction

This study addresses a series of questions concerning persistence and academic success among university students. What differentiates those students who continue in the university at which they first register from those who leave within one or two years? What characterizes those students who have a high quality educational experience as measured by accumulated credits and cumulative grade averages? What distinguishes those who complete a university degree from those who fail to do so within a reasonable span of time? Our data also allow us to investigate the above questions and to begin an exploration of some factors that may contribute to persistence such as: gender, type of program in which the student is enrolled, mother tongue, citizenship status, proximity to university, high school grade averages, characteristics of the student's high school, and socio-economic characteristics of the neighbourhood in which the student's family resides.

We use administrative data that have been collected on students in four Ontario universities and linked with information on the high school performance of the student, the neighbourhood in which the student lived during high school, and the high school which the student attended. The linked data sets are referred to collectively in this report as the persistence dataset. The university administrative data are from four reasonably representative Ontario universities, two of the medical/doctoral variety, one comprehensive university, and one primarily undergraduate university.[1] Jointly the four universities account for almost 30% of the student university population in Ontario.

This report is divided into the following sections: review of the literature; description of the data development process; descriptive statistics; multivariate analysis; and a summary of the findings and policy implications.

Review of the Literature

Mueller (2008a,b) and Parkin and Baldwin (2009) have recently completed very helpful reviews of the literature on Canadian postsecondary education (PSE) to which the interested reader is referred for a detailed discussion. Mueller confirms that persistence in post-secondary education is much less studied than is access and that the distinction is important. For example, post-secondary attendance rates in the US have risen much faster since 1970 than have completion rates Mueller (2008b). Persistence in post-secondary education is even less studied in Canada than in the U.S. but a number of useful papers on persistence have been published recently most commonly using data from the Youth in Transition Survey (YITS). Parkin and Baldwin (2009) discuss many of these studies in detail. The YITS data are indeed valuable but, as Finnie and Qiu (2009) observe, suffer from the problems commonly found in surveys: selection and response bias, limited sample size, limited information about high school performance and a lack of institutional detail.

[1] These designations are taken from a classification system for Canadian universities created by Maclean's magazine.

Finnie and Qiu (2008) used the YITS to illustrate the complex PSE pathways that Canadian students follow. For example, Finnie and Qiu reported that only 56.5 percent of college students and 52.1 percent of university students complete a degree or diploma at their initial PSE institution after five years of continuous enrolment. These completion rates rise to 73.1 and 69.4 percent, respectively, when one includes students who switched institutions and those who left school temporarily. Total persistence rates, which also take into account those who are still in PSE, push the graduation rates to 82 (college) and 89.8 (university) percent. Finnie and Qiu (2009) in a more recent study use system-wide administrative data for the Atlantic provinces to make the same point. They avoid some of the limitations of survey data by going to this administrative data source but are quite limited in the co-variates they are able to study.

In an early report with YITS data, Bowlby and McMullen (2002) found greater persistence among YITS students who had better financial resources, namely, parental income and scholarships. More recently, Martinello (2008) reports that parental education in the YITS was not correlated with successful completion of a first program (conditional on entry) but was correlated with students' ability to find and start second programs if they did not complete their first. Surprisingly, gender was not related to persistence among university students but male college students were much less likely to complete their first program. Johnson (2008) found little evidence in the YITS that either a higher level of tuition or a change in tuition alters the probability that a Canadian youth, once in university, leaves university without obtaining his or her degree. Johnson's study, however, took into account neither differences by field of study, across which tuition can vary greatly in some provinces, nor the impact of financial aid on the net cost of attending university. In a similar vein, Chemin (2009) used the YITS to assess the effect of a 2001 Quebec reform that raised the level of student grants. Chemin found that this reform increased PSE participation rates in Quebec relative to other provinces during the four years following the reform but that graduation rates in Quebec did not rise. Mueller (2008a) reported in his literature review that there is some evidence that finances play a weak role in persistence but this comes mainly from surveys that only ask students about why they drop out or do not, as the case may be.[2]

Administrative data has historically been used more by institutions hoping to improve their retention of their own students. Grayson and Grayson (2003) review the evidence from earlier studies and compare American studies and the very few Canadian studies. Many of the US studies have focused on differences in race and ethnicity, a topic of less relevance here in Ontario. More recent studies using administrative data include a US-based study by Nora et al (2005), the Finnie and Qui (2009) study mentioned above, and the recent HEQCO study by Conrad and Morris (2010) which focuses on how institutions might better analyze their retention issues. None of these studies, however, take the step of linking university administrative data with data for both high school and neighbourhood characteristics as we do here. Hence, the current study provides a major step forward in the use of university administrative data for research purposes.

[2]The Ontario Ministry of Training, Colleges and Universities reports that university degree completion rates within seven years of initial registration rose from 73.2% for the entry cohort of 1991 to 79.7% for entry cohort of 2002 (Ontario Ministry of Training, Colleges and Universities 2010).

Data and Measures

We make use of the following four administrative data sources: (1) administrative student records from four Ontario universities for varying periods from 1994-2006; (2) application data from the Ontario University Application Centre (OUAC) from 1994-2006; (3) characteristics of the neighbourhood from which the student applied to university as found in Enumeration and Dissemination Areas from the 1996, 2001 and 2006 Censuses; and (4) characteristics of the high school from which the student graduated as provided by the Ontario Ministry of Education for the years 2000 through 2003.

Measures of Academic Progress in University

We originally requested student-level information from all Ontario universities and four provided us with data for varying time periods. A detailed description of our data request is provided in the Data Appendix to this paper. Our request was limited to information for entering cohorts of '101 students', i.e., those entering a full time university degree program directly from an Ontario high school. These are the students for whom we already had application information from the Ontario University Application Centre (OUAC).[3] Furthermore, such students comprise approximately 90% of all entering students at Ontario universities (Dooley, Payne and Robb 2010). The earliest entering cohort for which we requested information was 1994 which is the earliest year for which we have OUAC data. The data provided by the four universities were quite similar in most regards save for two aspects. First, one university was only able to provide data starting in 1999.[4] Second, only two universities were able to provide scholarship data.

Excluded Records

We received a total of 172,143 records from the four universities. It proved convenient for some universities to provide us with records for some entrants which we did not request, e. g., pre-1994, Level 2 entry, part-time students, students not entering directly from high school, non-September entrants, and students from outside Ontario. We excluded such students from the sample used for this paper. We also removed a small number of otherwise acceptable students for the following reasons: the student's age was less than 15 or greater than 20 (we did not expect these ages); the student's postal code in the OUAC application data was not an Ontario postal code or could not be matched to an Enumeration/Dissemination Area in Ontario; a key variable, such as birth date or entry level, had a missing value; or the student had a record in two of our universities. When there were multiple records and the records had different entry years, we kept only the record with the earlier entry year. When there were multiple records and the records had the same entry year, we kept the record for the university at which the student enrolled for more than one year. See the Data Appendix for the specific number of students

---

[3] Our choice of these students also had to do with selecting a relatively homogeneous group of students .

[4] The shortened data period was due to the fact that historical administrative data was more complicated and expensive for that institution to retrieve.

excluded for each of the above reasons. Our final research data set contains records for 128,166 students.

Allocation of Students to Programs

We asked the universities to classify the students by academic program. We did so in recognition of the fact that some degree of disaggregation by program is desirable given the differences among faculties in entrance standards and academic programs. The numbers of students were sufficient to allow disaggregation into the following four categories: Arts, Science, Business and Engineering. Hence, we reallocated all students to one of these four entry programs. This process involved the reassignment, for example, of students in Kinesiology to Science and students in Music to Arts. Table 1 provides definitions and sample means or proportions of all the variables which we have used in our analysis. The first four rows indicate the distribution of our sample by entry program. Arts and Sciences account for three-quarters of our students and the professional programs, Engineering and Business, account for the remaining quarter.

Table 1
Definition and Sample Means or Proportions of Variables

Variable Name	Definition (“Year” refers to the twelve-month period following September entry)	Sample Mean or Proportion
Panel A: University Administrative Variables		
Arts Entry Program	Equal to 1 if enrolled in an Arts program in Year 1.	41%
Science Entry Program	Equal to 1 if enrolled in a Science program in Year 1.	35%
Business Entry Program	Equal to 1 if enrolled in a Business program in Year 1	12%
Engineering Entry Program	Equal to 1 if enrolled in a Engineering program in Year 1.	12%
Cumulative Credits Passed Year 1	Cumulative credits passed at the end of year 1 among all students.	4.9
Cumulative Credits Passed Years 1 & 2	Cumulative credits passed at the end of year 2 among students observed for 2 or more years.	9.2
Cumulative GPA Year 1	Cumulative grade point average (0-100) at the end of year 1.	71
Cumulative GPA Years 1 & 2	Cumulative grade point average (0-100) at the end of year 2 among students observed for 2 or more years.	72
% Departed During Year 1	Proportion of students for whom we observe only missing values in the second calendar year after entry among students observed for 2 or more years.	8%
% Departed During Years 1 & 2	Proportion of students for whom we observe only missing values in the third calendar year after entry among students observed for 3 or more years.	13%
% with Degree	Proportion of students for whom we observe a degree earned at the	45%

after 4 Years	end of 4 years after entry among students that we observe for 4 or more years.	
%with Degree after 5 Years	Proportion of students for whom we observe a degree earned at the end of 5 years after entry among students that we observe for 5 or more years.	74%
% with Degree after 6 Years	Proportion of students for whom we observe a degree earned at the end of 6 years after entry among students that we observe for 6 or more years.	80%
Panel B: Ontario Universities Application Centre Variables		
Female	Equal to 1 if the student is female.	57%
English Mother Tongue	Equal to 1 if the student's mother tongue is English.	85%
Canadian Citizen	Equal to 1 if the student is a Canadian citizen.	93%
Distance to Campus	Dummy=1 if home is more than 50 km from campus.	6%
Age at Entry	Age (in months) at entry to university	222
Average HS Grade	Average grade in the student's best six university or mixed HS courses.	83%
Average HS Grade< 75	Equal to 1 if the average grade is less than 75 (mostly greater than 70).	9%
Average HS Grade >=75 and <80	Equal to 1 if the average grade is equal to or greater than 75 and less than 80.	18%
Average HS Grade >=80 and <85	Equal to 1 if the average grade is equal to or greater than 80 and less than 85.	31%
Average HS Grade >=85 and <90	Equal to 1 if the average grade is equal to or greater than 85 and less than 90.	25%
Average HS Grade >=90 and <95	Equal to 1 if the average grade is equal to or greater than 90 and less than 95.	14%
Average HS Grade >= 95	Equal to 1 if the average grade is equal to or greater than 95.	3%
All University Courses	Equal to 1 if best six HS courses are all university level.	66%
Panel C: Neighbourhood Variables from Census Dissemination (Enumeration) Areas		
Average Equivalent Income	Average neighbourhood divided by the square root of the neighbourhood average number of persons per household ($2001).	$45,504
Low Income	Equal to 1 if the student comes from a neighbourhood in the bottom tercile of the distribution (weighted by total population) of all neighbourhoods by average equivalent income.	18%
Middle Income	Equal to 1 if the student comes from a neighbourhood in the middle tercile of the distribution (weighted by total population) of all neighbourhoods by average equivalent income.	31%
High Income	Equal to 1 if the student comes from a neighbourhood in the top tercile of the distribution (weighted by total population) of all neighbourhoods by average equivalent income.	50%
% Bachelor's	Proportion of adults in the neighbourhood with a degree at the	21%

Degree	Bachelor's level or higher.	
% Lone Mother	Proportion of families in the neighbourhood headed by a lone mother.	11%
% English	Proportion of persons in the neighbourhood with English as mother tongue.	87%
% Recent Immigrant	Proportion of persons in the neighbourhood immigrated since 1981.	13%
% Unemployed	Proportion unemployed of adults in the neighbourhood.	7%
	Panel D: Ministry of Education High School Variables	
% of High EQAO Scores in Bottom Tercile	Proportion of High EQAO Scores (3 or 4) in the High School is in the Bottom Tercile of all High Schools with OUAC Applicants.	22%
% of High EQAO Scores in Middle Tercile	Proportion of High EQAO Scores (3 or 4) in the High School is in the Middle Tercile of all High Schools with OUAC Applicants	32%
% of High EQAO Scores in Top Tercile	Proportion of High EQAO Scores (3 or 4) in the High School is in the Top Tercile of all High Schools with OUAC Applicants.	46%
Missing EQAO Scores	Equal to 1 if high school is missing EQAO scores.	8%
Distance to University	Distance (km) from high school to nearest university.	22 km.
Distance to College	Distance (km) from high school to nearest college.	11 km.
Private	High school is private (not publicly funded).	7%
Public, English	High school is public and English.	74.9%
Public, Francophone	High school is public and Francophone.	0.2%
Catholic, English	High school is Catholic and English.	24.5%
Catholic, Francophone	High school is Catholic and Francophone.	0.4%
Rural	High school is in rural area.	16%
% at High School with Total Enrolment in Bottom Tercile	Proportion of Students at a High School with Total Enrolment in the Bottom Tercile of all High Schools with OUAC Applicants.	24%
% at High School with Total Enrolment in Middle Tercile	Proportion of Students at a High School with Total Enrolment in the Middle Tercile of all High Schools with OUAC Applicants.	34%
% at High School with Total Enrolment in Top Tercile	Proportion of Students at a High School with Total Enrolment in the Top Tercile of all High Schools with OUAC Applicants.	44%
	Panel E: Entry Years	
1994 Entry Year	Equal to 1 if enrolled in Year 1 in Fall of 1994.	5%
1995 Entry Year	Equal to 1 if enrolled in Year 1 in Fall of 1995.	5%
1996 Entry Year	Equal to 1 if enrolled in Year 1 in Fall of 1996.	5%
1997 Entry Year	Equal to 1 if enrolled in Year 1 in Fall of 1997.	6%
1998 Entry Year	Equal to 1 if enrolled in Year 1 in Fall of 1998.	6%
1999 Entry Year	Equal to 1 if enrolled in Year 1 in Fall of 1999.	10%

2000 Entry Year	Equal to 1 if enrolled in Year 1 in Fall of 2000.	10%
2001 Entry Year	Equal to 1 if enrolled in Year 1 in Fall of 2001.	10%
2002 Entry Year	Equal to 1 if enrolled in Year 1 in Fall of 2002.	11%
2003 Entry Year	Equal to 1 if enrolled in Year 1 in Fall of 2003.	14%
2004 Entry Year	Equal to 1 if enrolled in Year 1 in Fall of 2004.	9%
2005 Entry Year	Equal to 1 if enrolled in Year 1 in Fall of 2005.	6%
2006 Entry Year	Equal to 1 if enrolled in Year 1 in Fall of 2006.	4%

Time Frame for Analysis

The universities reported courses and grades by academic term (fall, winter and summer). One challenge for measuring academic progress is posed by coop programs. All of our universities have coop programs though the degree of student participation varies by institution. Students in coop programs typically are less likely than non-coop students to be enrolled in fall and winter terms (save for the fall of year 1) and are more likely to be enrolled in summer terms. Put differently, course loads are more evenly spread throughout the calendar year for coop students than for non-coop students. Students in coop programs still do take vacation terms and, as a result, progress towards a degree is typically slower than for non-coop students. Unfortunately, the data do not directly indicate which students are in coop programs or which students have no credits in a given term due to being on a coop term. In our multivariate analysis, we control for this to some extent by including control variables for university and program. A second challenge is posed by full-year (two-term) courses which are most common in the first year. Such courses do not show up as passed or failed until the winter term and, as such, can give the mistaken appearance of a light course load when one looks at courses completed in the fall term.

We incorporate the above considerations into our measures of academic progress in two ways. First, we measure such indicators as credits completed and cumulative grade averages on a calendar year basis rather than on an academic term basis. Second, when measuring degree completion rates, we use a wide data window so as to allow for the slower pace of coop students.

Measures of Academic Progress

Credits completed towards degree

Panel A of Table 1 reports the measures of academic persistence and success, the first two of which are the cumulative credits passed at the end of 1 and 2 calendar years after entry. We believe that these are the best available measures of the student's quantitative progress towards a degree. By "year", we mean the twelve-month period following the September entry, that is, to the end of the summer term. We use a credit measurement system under which 0.5 credits is given for a one-term course and, hence, 5 credits would be the most common one-year, full-time academic load. The sample mean of credits completed after one year is 4.8 and the sample mean after two years is 9.2. Note that these two means are based on different sample sizes. This reflects our decision to use all available data in our analysis. We have only one year of data for the most recent cohorts of students in our samples, two years of the data for the penultimate cohorts and so on. Hence, we observe more students for one than for two years. The relationships between the one-year and two-year measures and the characteristics of the individuals, neighbourhoods and high schools in our sample were quite similar. Hence, we focus on the results for cumulative credits passed after two years in the analysis below.

Cumulative Grade Averages

We believe that the best available measures of the student's qualitative progress towards a degree is the cumulative grade average at the end of 1 and 2 calendar years after entry. Some of our universities use a 0-100 grading system and some use 0-12 system. We transform all of the systems to a system that uses a 0-100 grade range. The biggest challenge we faced when transforming the grade averages was with respect to the treatment of a failing grade (F). The method of dealing with this issue varies not only across universities but in some cases also across Faculties within a university. We took a number of steps to assess the impact of grading differences. One was to conduct separate analyses among students within each university. A second was to compare students across universities using measures which are less sensitive to the above issues, e.g., having a grade average of at least A or at least B. A third was to convert all grades to a 0-100 point system using various assumptions to make the necessary conversions.[5] Each of these strategies for measuring grades led to conclusions which are very similar to those reported in the tables and figures below. The sample means when we do convert to the 0-100 point system (in Table 1) are 71 after one year and 72 after two years. The standard deviations (not shown) are both equal to 10.

Departures

Universities and colleges are all very interested in the problem of dropouts and retention. Hence, we also measure whether the student has departed from the university within year 1 and within years 1 and 2. We use the term "departure" rather than "dropout" because the former more accurately reflects what one can infer from our data. We do not know the destination of students who cease to register for courses in our four universities. Some departures undoubtedly are voluntary transfers to other universities or colleges or temporary absences from PSE as documented by Finnie and Qiu (2008). Indeed the only means we have of inferring a departure is from courses passed and failed term by term. We note, however, that those students who do depart have lower than average grades and lower than average cumulative credits completed. We count as a "departure during the first year" any student for whom we have no courses passed or failed in the fall, winter or summer of the second academic year. (The number of "departures during the first year" is only slightly lower if we count the students for whom we observe no grades (pass or fail) for courses in the fall, winter or summer of the second and third academic years.) We count as a "departure during the first two years" any

[5] We converted in two completely different ways and the results differed only slightly. One way was to convert the 0-12 numbers to percentages while preserving the corresponding letter grade interpretations of the two systems. That is, we ensured that an A was still an A, a B was still a B, a failure was still a failure, and so on. The second method was to create a standard normal distribution of grades for each university and then scale each university's grades so that they had the same mean and standard deviation as was found in the institutions that originally graded on a 0-100 scale. Further details are available on request.

student for whom we have no grade for courses in the fall, winter or summer of the third academic year. For each of the departure measures, we study only those students belonging to cohorts for whom we have data for the appropriate years of study (e.g. two years of data for the one year departure measure and three years of data for two year departure measure) We also make sure not to count a student who has earned an early degree as a "departure". Table 1 indicates that the departure rate during the first year is 8% and during the first two years is 13%. The difference between these two figures is the 5% of entering students who depart during year two.

Degree Completion

We believe that the best single measure of academic success that is available from our data is whether or not the student has completed a degree within a given time frame. Our final measure of persistence is degree completion. We have information on whether or not a degree is earned, the type of degree earned (honours or general) and the time taken to earn the degree. Most of the degrees in our data would typically take three or four years to complete if pursued on a full-time basis but there are also five year joint degree programs. More importantly, there are many students in our data who take more than the minimum numbers of full-time years needed to complete a degree program due to coop terms, program switches, periods of part-time study, academic terms abroad, etc. Only 45% of the students in our sample complete any degree within four years. Students completing a degree within four years are disproportionately from Arts and Science programs and from those universities at which coop programs are less common. The proportions completing a degree within five or six years are 74% and 80% respectively. Such longer data windows also yield a very different pattern of degree completion rates by program and university.

We estimated our multivariate models using each of these measures of degree completion, namely, whether or not a student had completed a degree with four, five or six years of entry. The longer data windows reveal, we believe, a more accurate picture of completion rates by program. In addition, we found that the coefficients for individual, neighbourhood and high school characteristics are quite similar across regressions with different measures of degree completion. For these reasons, we focus in this report on the regression estimates in which the dependent variable is whether or not the student earned any degree within six years after entry. We found this also to be a measure used in other studies of academic success in university (for example see Nora, Barlow and Crisp 2005). [6]

[6] Our analysis of the data indicated that the distinction between a "General" degree and an "Honours" degree was not that informative. Many "General" degree programs take four full-time years just like "Honours" degrees. Furthermore, some non-Honours programs, such as Engineering, make academic demands of the student that are at least as extensive as Honours degrees in other disciplines.

Individual, Neighbourhood and High School Characteristics

Individual Characteristics.

Panel B of Table 1 provides definitions and sample means or proportions of the individual, neighbourhood and school characteristics that we link to our measures of academic progress. The OUAC data confirmed that 55% of the students in our sample are females, 85% have English as mother tongue, 93% are Canadian citizens, and 94% have a home residence within 50 kilometres of campus. The students' mean age at registration is 18.5 years. OUAC data also provide the average grade in the student's best six university level courses in high school.[7] Table 1 indicates that the sample mean of this variable is 83. Seventeen percent of students have an average of 90 to 100 and 27% have an average of less than 80.

Neighbourhood Characteristics

The residential postal code in the OUAC data allows us to link the student's record to the data for Enumeration Areas (1996) and Dissemination Areas (2001, 2006) in the census. The Data Appendix provides a detailed description of how this was carried out. Our key income measure is "average equivalent income" which is average neighbourhood income divided by the square root of the neighbourhood average number of persons per household. The mean value of "average equivalent income" in our data is $45,504 (2001 dollars). This measure is analogous to the most commonly used strategy of adjusting individual household income for household size which is to divide household income by the square root of the number of person in the household.[8]

We also calculated the distribution of average equivalent income across all Enumeration (1996) and Dissemination (2001, 2006) Areas weighted by total population. As indicated in Panel C of Table 1: 18% of the students from our four universities are from neighbourhoods in the bottom tercile (one-third) of this distribution, 31% are from the middle tercile and 50% from the top tercile. These values reflect that fact that university students come disproportionately from more affluent neighbourhoods.

Table 1 also describes five other neighbourhood measures that we use. Our students come from neighbourhoods in which, on average, 21% of the adults have a Bachelor's Degree or more, 11% of the families are headed by a lone mother, 87% of persons have English for a mother tongue, 13% of persons have immigrated to Canada since 1981, and 7% of adults are unemployed.

High School Characteristics

The next set of measures in Panel D of Table 1 are the characteristics of high schools derived

---

[7] These are six OAC courses for students who took Grade 13 and six U or U/M courses for later cohorts.
[8] Standardizing households of different sizes by use of an 'equivalence scale' is now quite common. Four people do not need four refrigerators. The square root scale is one of the simplest and most commonly employed.

from data provided by the Ministry of Education and averaged over the years 2000 through 2003 (our data are incomplete over other years). One measure we use is the proportion of students in the high school taking the Grade 9 Academic EQAO Exam that received a "high score" (3 or 4). Using this measure, we ordered all high schools in the OUAC data and determined the cutoffs for the bottom, middle and top terciles.[9] These cutoffs are 61% for the bottom tercile and 71% for the top tercile.[10] Table 1 reveals that 22% of the students in the persistence data are from high schools in the bottom tercile of this frequency distribution, 32% of students are from school in the middle tercile and 46% are from schools in the top tercile. As one might expect, students at the four universities in our data set come from high schools in which students do better than average (across Ontario) on the Grade 9 EQAO exams.

Table 1 also reveals that 8% of students in the persistence data set are from high schools for which EQAO scores are missing. The average student in the persistence data set attended a high school that is 34 kilometres from the nearest university and 24 kilometres from the nearest college. Seven per cent of students come from private high schools, 74.9% from English public schools, 0.2% from Francophone public schools, 24.5% from English Catholic schools, 0.4% from Francophone Catholic schools, and 16% from rural high schools.[11] Finally, we grouped all high schools in the OUAC data by total enrollment and determined the cutoffs for the bottom, middle and top terciles of school size. Although it is the schools that are grouped , we do so by 'weighting' by student enrollments so that 1/3 of the students (not high schools) are in each tercile. These cutoffs are 924 students for the bottom tercile and 1310 students for the top tercile. Table 1 reveals that 24% of the students in the persistence data are from high schools in the bottom tercile of this frequency distribution, 34% of students are from school in the middle tercile and 44% are from schools in the top tercile. Students at the four universities in our data set come from high schools that are larger than average in terms of total enrolment.

Panel E of Table 1 portrays the distribution of students in our data by year of entry to university. Changes in this proportion over time reflects both changing cohort size, especially the "double cohort", and the fact that different universities provided data for differing periods of time (see Section 3.1 above). In our regressions, we use both binary variables for year of registration and age at registration to control for differences among cohorts including the elimination of the OAC year and the associated curriculum reform.

Summary Statistics

Table 2 provides comparisons between students from the four universities in our persistence data and all students in our OUAC data, that is, all applicants to all universities in Ontario applying directly from Ontario High Schools – known as '101' applicants named after the form completed by such students. We focus on 1999 through 2004 which are the years for which we

[9] Approximately 90% of high schools with Grade 12 in Ontario in 2000 appear in our OUAC data file, that is, have at least one student who applies to an Ontario university.
[10] These cutoff values are very similar in the enrolment weighted and unweighted frequency distributions which implies that performance on the Grade 9 EQAO test and school size are not strongly correlated.
[11] There is a large overlap between private schools and missing EQAO scores. Nine-five percent (95%) of private high schools have missing EQAO scores and 77% of the schools with missing EQAO scores are private.

have data from each of our four universities. This subset of our total sample contains 83,496 students from our four universities represent 20% of all the 101 OUAC applicants and 28% of all 101 OUAC registrants over the same time period. The proportion female of all three groups are similar. The persistence students are slightly more likely than the other two groups to have a mother tongue that is English and to be Canadian citizens. Students in the persistence data are also bit younger and have higher grades both in the average of their best six university courses and in the core Grade 12/OAC English.[12] We believe that Table 2 indicates that the students in our persistence data set are reasonably similar to all students in Ontario universities.

Table 2
Students Applying and Registering Directly From An Ontario High School: 1999-2004*

	Total OUAC Applicants	Total Registrants in Ontario Universities	Persistence Data
Number of Students	426,594	298,697	83,496
% of OUAC Applicants	100%	70%	20%
% of Ontario University Registrants	143%	100%	28%
% Female	55.9%	57.9%	56.5%
% English mother tongue	78.7%	79.8%	84.6%
% Canadian citizen	89.9%	91.5%	93.1%
Mean Age at Entry	19.4	19.1	18.5
Mean grade in best six Grade 12/OAC courses	78.5	81.3	83.4
Mean Grade OAC/Grade 12 core English course	76.4	78.5	80.1

*The years 1999 through 2004 for which have data for entering classes from all four universities in the persistence data.

Figures 1a through 1d contain the time series for four measures of academic progress at each of our universities.[13] The four measures of progress are mean credits completed by the end of Year 2, mean cumulative grade average at the end of Year 2, the proportion of entrants departing during the first two years, and the proportion of entrants completing a degree within six years of entry. We draw three basic conclusions from these figures. First, there are only minor differences among our four universities and the rank order changes over time. Second, these measures have generally been quite stable over time at each of our universities. There is no indication that rising tuition (and other changes in the policy and academic environment) have led to lower (or higher) levels of persistence and/or degree completion. Third, the figures do reveal some minor improvements in those outcomes which can be measured for the “double cohort” that entered in 2003, namely, credits completed, grade averages and departure rates. This may reflect the greater selectivity which universities were able to exercise at that entry point. The average grade in the best six Grade 12/OAC courses rose from 83 among 2002 entrants to 85 among 2003 entrants and then fell back to 83 in 2004.

[12] We chose to present the core English grade because we have this mark for virtually all students in our data. This is not the case with Grade 12/OAC math courses or indeed any other course.
[13] We cannot identify the universities for reasons of confidentiality.

Figure 1: Means of Persistence Measures by Entry Year and University

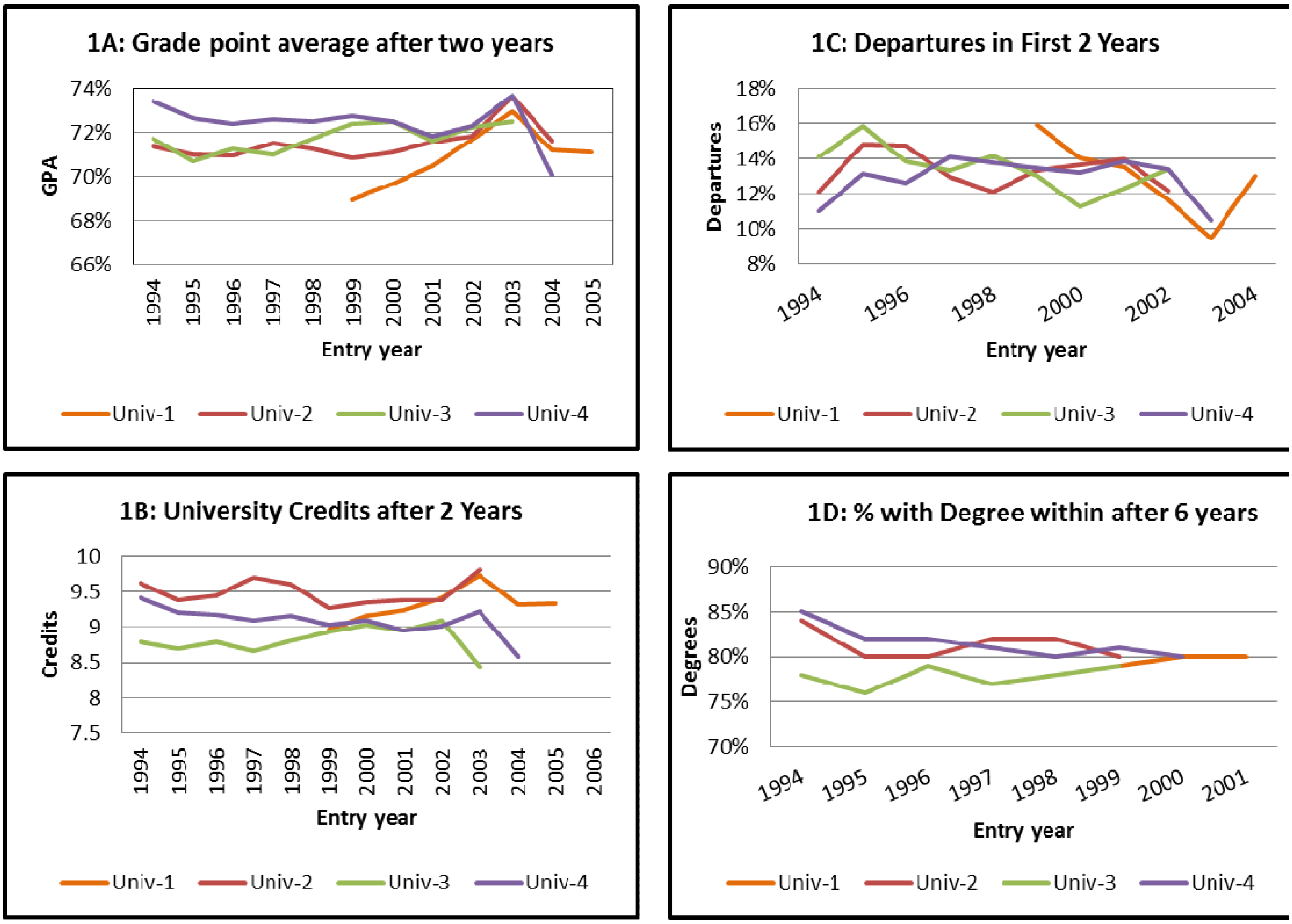


Table 3 provides various summary statistics for our measures of academic persistence. The top row provides the sample means and proportions for all of the students in the persistence sample (for whom we observe the number of years required to compute the measure.)[14] Column 2 indicates that, among our total sample of 128,166 students whom we observe for one or more years, the mean credits passed after 1 calendar year (at the end of the summer term following entry) was 4.9 which is just short of the most common full time load of 5. This high mean is accounted for in part by the fact that a substantial proportion of students take more than 5 credits in their first year especially in engineering programs and that some students enrol in

[14] Note that the statement "observed for 2 or more years" means "we have 2 or more years of data". It does not mean "enrolled for 2 or more years". Some of the students "observed for 2 or more years" left the university during their first year.

summer school as well as in the fall and winter terms. Column 3 indicates that, among the 113,271 students that we observe for two years, the mean credits passed after 2 calendar years (at the end of the second summer term following entry) was 9.2. Columns 4 and 5 indicate that the mean GPA is 71% after one year and 72% after two years.

Table 3 Summary Statistics for Measures of Persistence by Entry Program, Gender, High School GPA and Neighbourhood Average Equivalent Income

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mean Credits after 1 year	Mean Credits after 2 years	Grade Point after 1 year	Grade Point after 2 years	% Departed during Year 1	% Departed during Years 1 and 2	% Any Degree within 6 years
Total	4.9	9.2	71	72	8%	13%	80%
By Entry Program							
Arts	4.6	8.7	70	70	10%	17%	75%
Science	4.9	9.1	72	73	7%	12%	81%
Business	4.9	9.4	73	73	5%	9%	88%
Engineering	5.4	9.7	71	72	6%	9%	82%
By Gender							
Male	4.8	8.9	71	71	8%	13%	77%
Female	4.9	9.2	72	72	8%	13%	82%
By High School Grade Point Average							
<75	4.1	7.4	63	63	19%	28%	58%
=>75 and <80	4.5	8.4	66	67	12%	19%	72%
=>80 and <85	4.9	9.1	70	70	7%	12%	81%
=>85 and <90	5.1	9.5	74	74	5%	9%	87%
=>90 and <95	5.2	9.7	80	79	3%	6%	91%
=>95	5.4	9.9	87	86	2%	5%	94%
By Neighborhood Average Equivalent Income							
Bottom Tercile	4.8	8.9	71	71	10%	15%	77%
Middle Tercile	4.9	9	71	72	8%	13%	80%
Top Tercile	4.9	9.1	72	72	7%	12%	81%
By Proportion of Students Achieving High Score (3 or 4) on Academic EQAO Test							
Bottom Tercile	4.8	9	70	71	9%	14%	78%
Middle Tercile	4.8	9.2	71	72	8%	13%	80%
Top Tercile	4.9	9.2	72	72	7%	12%	81%
Number of Observations	128166	113,271	128166	113,271	113,271	97,558	55,574
Number of Years Observed	1 or more years	2 or more years	1 or more years	2 or more years	2 or more years	3 or more years	6 or more years

In Section 3.5 above, we defined a “departure during Year 1” as a student for whom we have no courses passed or failed in the fall, winter or summer of the second academic year. Column 6 shows that 8% of students whom we observed for two or more years departed during their first year. We defined a “departure during Years 1 and 2” as a student for whom we have no courses passed or failed in the fall, winter or summer of the third academic year. Column 7 shows that the proportion of such departures among all students whom we observe for three or more years is 13%. Column 8 indicates that, among those students observed for at least six years in our data, 80% have earned a degree within 6 years.

The remaining rows of Table 3 show the variation in these outcome measures by different characteristics. The second panel shows that the students who enter the professional programs, Business and Engineering, have more credits completed, higher grades, lower departure rates and higher degree completion rates than do students who enter the other programs, especially Arts. The third panel of Table 3 shows the differences by gender. Females have better outcomes than males but these differences are small except for that fact that women have noticeably higher degree completion rates. Our multivariate analysis below shows that females perform better than males on all measures when one controls for other important gender differences, especially entry program.

The fourth panel of Table 3 illustrates a key finding of our report which is the very strong link between our university outcomes and high school grade point averages. Students in the lowest high school grade category (less than 75% average) and those in the highest high school grade category (95% or better) have university GPA differences of about the same magnitude, 25 percentage points. The departure rates in years 1 and 2 of these same two groups of students differ by 23 percentage points and the degree completion rates differ by 36 percentage points. Save in the case of university grades, there also appears to be non-linearity, namely there are bigger differences in outcomes between groups of students at the bottom end of the grade distribution than at the top.

The penultimate panel of Table 3 contains the differences in the outcome measures by neighbourhood income tercile which we consider to be relatively minor. The differences in credits completed and grade averages between students from low income neighbourhoods and students from high income neighbourhoods are also very small. The difference in departure rates is three percentage points and that in degree completion rates is four percentage points. Students from middle income neighbourhoods always occupy a middle ground.

The final panel of Table 3 contains the differences in the university outcomes by one measure of the performance of the high school from which the student graduated, namely the proportion of students taking the Grade 9 Academic EQAO test that achieved a high score (3 or 4). As with differences by neighbourhood income, we would characterize these differences in university outcomes as modest. Students from high schools in bottom and top percentiles of this frequency distribution differ little in credits completed and grade averages. The difference in departure rates is two percentage points and the difference in the degree completion rate is three percentage points.

Multivariate Analysis

We estimated a wide variety of regression models using a large number of academic outcomes including credits completed and cumulative grade average at the end of Year 1 and Year 2, the departures during Year 1 and during Year 2, and degrees completed within four, five and six years after entry. The estimated coefficients were quite robust across different outcome measures of the same type (e.g. degrees earned after different lengths of time).[15] Accordingly, we have chosen to present only the results for the four persistence measures the means of

[15] The major exception was the aforementioned finding that the coefficients for academic program differed considerably in the regression for degrees completed within four, five and six years after entry, presumably because programs have different normal lengths of study.

which were presented in Figure 1, namely, credits completed by the end of Year 2, cumulative grade average at the end of Year 2, departures during Years 1 and 2, and degrees completed within six years after entry. We estimated the regressions for these four outcomes in several different forms. Two of these are in Tables 4 and 5. Each regression contains a dummy variable for each university the estimates of which are not included in this report due to confidentiality considerations. The regression in Table 4 also contains a series of individual, neighbourhood and high school characteristics listed in Table 1 but there are no interactions of any variables. The regressions in Table 5 are the same as those in Table 4 save for the fact that we have included interactions between high school grade average categories and each of the variables for entry program, gender, neighbourhood income and high school performance on the Grade 9 EQAO test.[16]

[16] Given that only two universities provided data on scholarships and that entry scholarship values at those two institutions are highly correlated with high school grade averages, we have opted not to include this variable in our analysis

Table 4 Regressions With No Interactions

	(1)	(2)	(3)	(4)	(5)
	Dependent Variable	Cumulative Grade Average After Two Years	Credits Passed After Two Years	Departed During Years 1 & 2	Completed Degree Within 6 Years
(1)	Science Entry Program	-0.965***	0.181***	-0.012***	0.010**
		(0.093)	(0.021)	(0.003)	(0.005)
(2)	Business Entry Program	-0.806***	0.274***	-0.037***	0.058***
		(0.104)	(0.030)	(0.004)	(0.006)
(3)	Engineering Entry Program	-4.051***	0.832***	-0.026***	0.008
		(0.132)	(0.043)	(0.004)	(0.007)
(4)	Female	0.849***	0.393***	-0.008***	0.047***
		(0.060)	(0.018)	(0.003)	(0.004)
(5)	English Mother Tongue	0.640***	-0.000	0.002	-0.007
		(0.122)	(0.030)	(0.004)	(0.006)
(6)	Canadian Citizen	0.733***	0.176**	-0.027***	0.039***
		(0.272)	(0.073)	(0.006)	(0.009)
(7)	Home is 50 km Or More From University	-0.801***	-0.186***	0.015***	-0.008
		(0.168)	(0.029)	(0.004)	(0.006)
(8)	Age at Entry (months)	-0.079***	-0.018***	0.001***	-0.003***
		(0.007)	(0.002)	(0.000)	(0.000)
(9)	Best Six Grade 12/13 Courses All University	1.056***	0.195***	-0.022***	0.083***
		(0.114)	(0.033)	(0.005)	(0.021)
(10)	HS Average Grade < 75	-7.627***	-1.650***	0.148***	-0.219***
		(0.101)	(0.040)	(0.006)	(0.008)
(11)	HS Average Grade =>75 and <80	-3.831***	-0.664***	0.057***	-0.087***
		(0.068)	(0.025)	(0.004)	(0.005)
(12)	HS Average Grade =>85 and <90	4.386***	0.361***	-0.032***	0.055***
		(0.061)	(0.020)	(0.003)	(0.004)
(13)	HS Average Grade =>90 and <95	10.094***	0.561***	-0.057***	0.102***
		(0.093)	(0.025)	(0.003)	(0.005)
(14)	HS Average Grade =>95	17.066***	0.682***	-0.069***	0.140***
		(0.163)	(0.039)	(0.005)	(0.007)
(15)	% High Scores on Grade 9 EQAO Test in Bottom Tercile	-1.086***	-0.124***	0.013***	-0.018***
		(0.206)	(0.036)	(0.004)	(0.006)
(16)	% High Scores on Grade 9 EQAO Test in Middle Tercile	-0.630***	-0.060*	0.006	-0.010
		(0.190)	(0.033)	(0.004)	(0.006)
(17)	No Grade 9 EQAO Test	1.602**	0.060	-0.006	0.020
		(0.658)	(0.122)	(0.017)	(0.022)
(18)	Distance of High School from Nearest University (km)	-0.002	0.000	-0.000	0.000
		(0.002)	(0.000)	(0.000)	(0.000)
(19)	Distance of High School from Nearest College (km)	0.020***	0.003***	-0.000***	0.000***
		(0.004)	(0.001)	(0.000)	(0.000)
(20)	Private High School	-2.901***	-0.347**	0.033*	-0.046*
		(0.786)	(0.154)	(0.019)	(0.025)
(21)	English Catholic High School	-1.015***	-0.074**	-0.005	0.009
		(0.176)	(0.033)	(0.004)	(0.006)
(22)	Francophone Public High School	1.322***	-0.134	0.042**	-0.034
		(0.456)	(0.158)	(0.020)	(0.034)
(23)	Francophone Catholic High School	-0.317	-0.142	0.010	0.003
		(0.629)	(0.132)	(0.015)	(0.023)
(24)	Rural High School	0.386*	0.056	-0.000	0.016**
		(0.211)	(0.041)	(0.006)	(0.007)
(25)	High School Enrolment in Bottom Tercile	-1.015***	-0.168***	0.020***	-0.019***
		(0.240)	(0.046)	(0.005)	(0.007)
(26)	High School Enrolment in Middle Tercile	-0.460**	-0.052*	0.011***	-0.007
		(0.193)	(0.031)	(0.004)	(0.005)
(27)	Low Income EA/DA	-0.159	-0.077*	0.010**	-0.020***
		(0.165)	(0.040)	(0.005)	(0.007)
(28)	Middle Income EA/DA	-0.002	0.003	0.002	-0.003
		(0.099)	(0.023)	(0.003)	(0.005)
(29)	% EA/DA Bachelor's Degree	0.018***	0.001	-0.000**	0.000**
		(0.006)	(0.001)	(0.000)	(0.000)
(30)	% EA/DA Lone Mother Families	-0.006	-0.004***	0.000	-0.001***
		(0.006)	(0.001)	(0.000)	(0.000)
(31)	% EA/DA English Mother Tongue	-0.019***	-0.002	-0.000	0.000
		(0.006)	(0.001)	(0.000)	(0.000)
(32)	% EA/DA Immigrant since 1981	-0.014**	-0.002	-0.000	0.000
		(0.006)	(0.001)	(0.000)	(0.000)
(33)	% EA/DA Unemployed	-0.018**	-0.003	0.000	0.000
		(0.009)	(0.002)	(0.000)	(0.001)
(34)	1994 Entry Year	0.000	0.000	-0.005	0.017
		(0.000)	(0.000)	(0.009)	(0.010)
(35)	1995 Entry Year	-0.347**	-0.083*	0.012	-0.002
		(0.135)	(0.047)	(0.009)	(0.011)
(36)	1996 Entry Year	-0.217	-0.024	0.002	0.008
		(0.145)	(0.049)	(0.009)	(0.010)
(37)	1997 Entry Year	0.059	0.026	-0.008	0.012
		(0.143)	(0.049)	(0.009)	(0.010)
(38)	1998 Entry Year	0.228	0.063	-0.005	0.012
		(0.143)	(0.046)	(0.009)	(0.010)
(39)	1999 Entry Year	0.250*	-0.058	-0.006	0.013
		(0.137)	(0.045)	(0.008)	(0.008)
(40)	2000 Entry Year	0.323**	0.022	-0.013*	0.011
		(0.150)	(0.048)	(0.008)	(0.009)
(41)	2001 Entry Year	0.443***	0.043	-0.012	0.000
		(0.155)	(0.048)	(0.007)	(0.000)
(42)	2002 Entry Year	0.394***	0.047	-0.011	0.000
		(0.151)	(0.043)	(0.008)	(0.000)
(43)	2003 Entry Year	0.228	-0.082*	-0.022***	0.000
		(0.161)	(0.046)	(0.007)	(0.000)
(44)	2004 Entry Year	-0.806***	-0.225***	0.000	0.000
		(0.212)	(0.055)	(0.000)	(0.000)
(45)	2005 Entry Year	-0.835***	-0.179***	0.000	0.000
		(0.232)	(0.060)	(0.000)	(0.000)
(46)	Constant	89.757***	13.320***	-0.129***	1.338***
		(1.474)	(0.416)	(0.049)	(0.081)
	Observations	113,271	113,407	97,558	55,574
	R-squared	0.380	0.113	0.042	0.073
	Sample mean or proportion	72	9.2	13%	80%

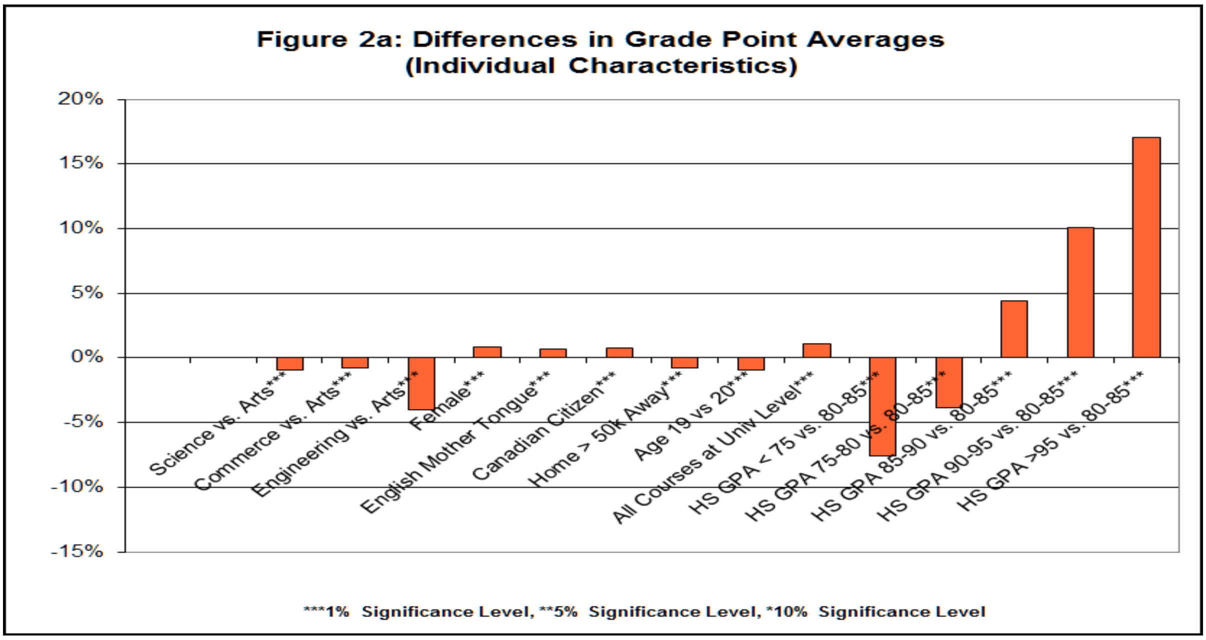
Clustered standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Constant (reference group) is for a male, Arts, non-English mother tongue, not a citizen, resides within 50 km, some non-University courses, GPA 80-85, high income, top tercile EQAO scores, 2006 entry. High school is publicly funded, English, public, Anglophone and urban. Each regression also contains a dummy variable for each university.

Figures 2 through 5 present bar charts of the coefficients from Table 4 for each of our four measures of academic success. For each figure, we separate the coefficient estimates into three groups: individual characteristics, high school characteristics and neighbourhood characteristics. In the case of binary independent variables, we just show the value of the coefficient. In the case of continuous independent variables, we show the coefficient multiplied by a representative change in the variable as indicated in the figure titles. In the case of the neighbourhood unemployment rate, for example, we show the impact of a one percentage point increase in this rate.

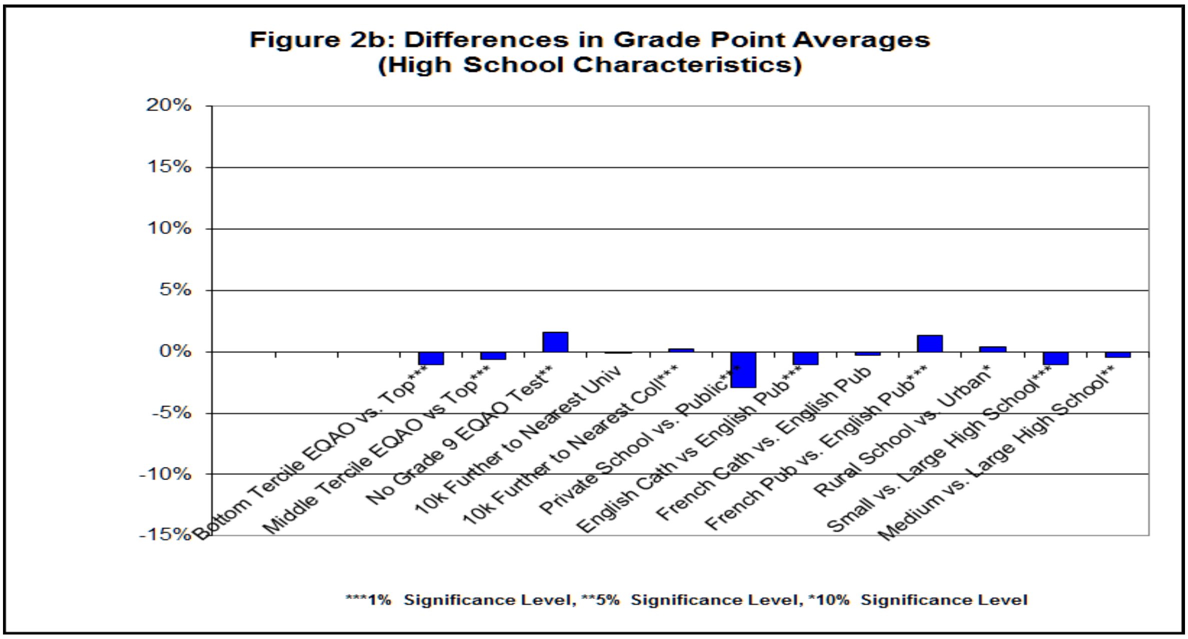
Figure 2 shows the coefficients from the first regression in Table 4 in which the dependent variable is the cumulative grade average two years after entry. Figure 2a charts the coefficients for individual characteristics. The key measure that stands out is the importance of the student's high school grade average. The strong link between university and high school grade averages is virtually the same as that noted above in Table 3 even after controlling for many other variables.[17] At the end of two years, students in the lowest high school grade category (less than 75) and those in the highest high school grade category (95-100) have university grade averages that differ by about 25 percentage points. The most noticeable difference among programs is that Engineering students have cumulative grade averages that are 3 to 4 percentage points lower than those of other students. The grade averages of females are about 1 percentage point higher than those of males.

Figure 2: Difference in GPA



The differences in university grades associated with the remaining individual characteristics in Figure 2a are often statistically significant but numerically small. Students have higher university grades that have English as a mother tongue, are Canadian citizens and have all university courses in their best six high school courses. "Commuters" do slightly better than those living more than 50 km from campus and entrants aged 20 do slightly worse than those aged 19 which may reflect a so-called "victory lap" phenomenon (students who take an extra year of high school to improve grades.)

Figure 2b contains the coefficients for high school characteristics in Table 4. The difference in cumulative grade averages between students from high schools with average EQAO scores in the bottom and top terciles is only 1 percentage point. Students from schools with no EQAO scores do slightly better.[18] All else equal, students from privately funded high schools have university grade averages that are 3 percentage points lower than those of students from publicly funded English schools. Students from English Catholic high schools do worse, and those from Francophone public schools do a bit better, than those from English public schools but these are small differences of 1 percentage point or less. The same is true of the superior performance of students from schools in rural areas.[19]



[18] For purposes of estimating these regressions, students from high schools with missing EQAO scores were included in the group of students from high schools with average academic EQAO scores in the middle tercile. We also estimated the regressions excluding students from high schools for which EQAO scores were missing and the resulting coefficients were very similar to those in Tables 4 and 5.
[19] It is important to remember that the differences in persistence between schools of different types estimated here condition on high school grades. Thus, the differences in persistence may reflect many factors including grading standards, curriculum, family background, etc. Our data do not allow us to distinguish among these possible causes.

Figure 2c contains the coefficients for neighbourhood characteristics in Table 4. The difference in cumulative grade averages between students from low income and high income neighbourhoods is less than 1 percentage point and statistically insignificant. The estimated impact of other neighbourhood characteristics is equally small even when statistically significant.

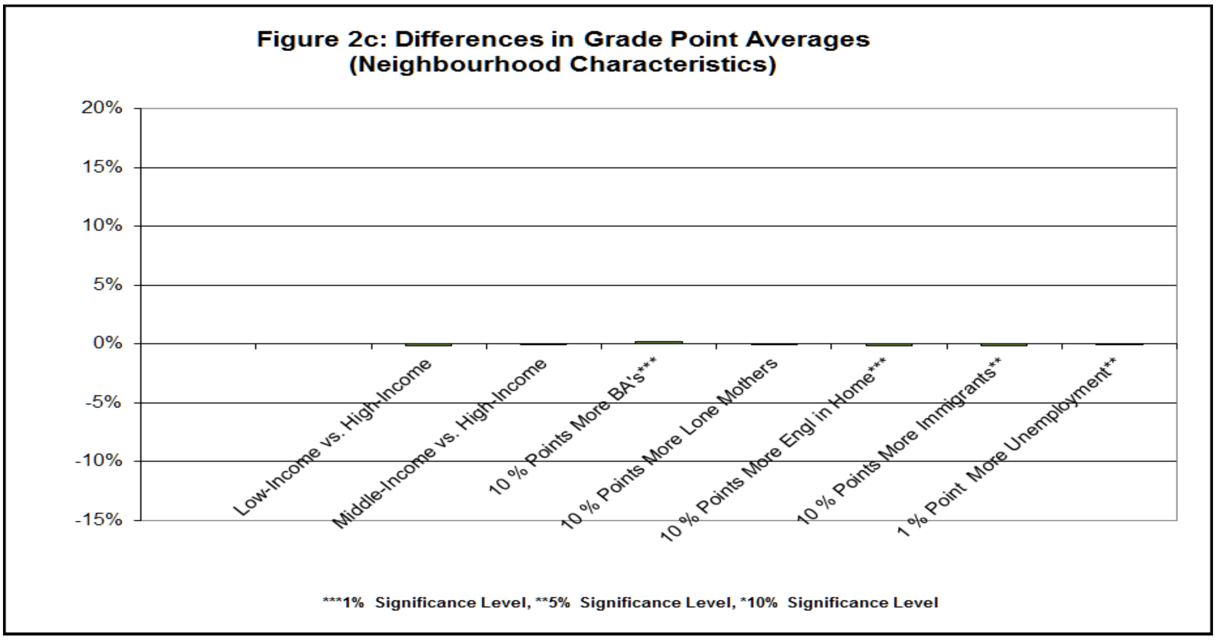
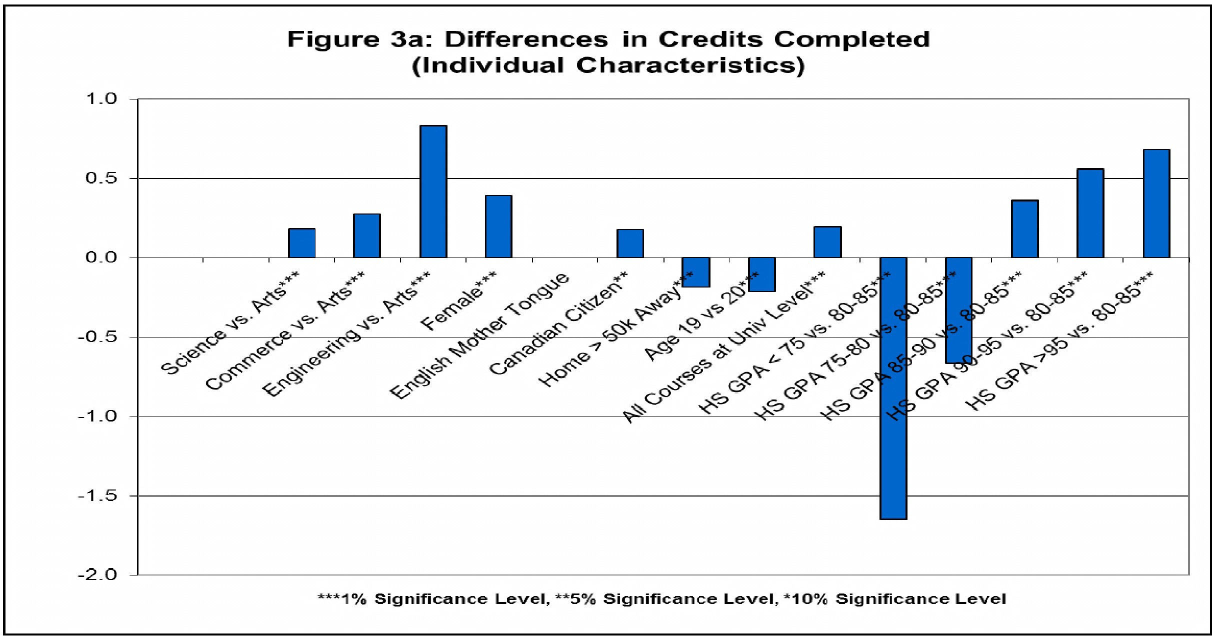
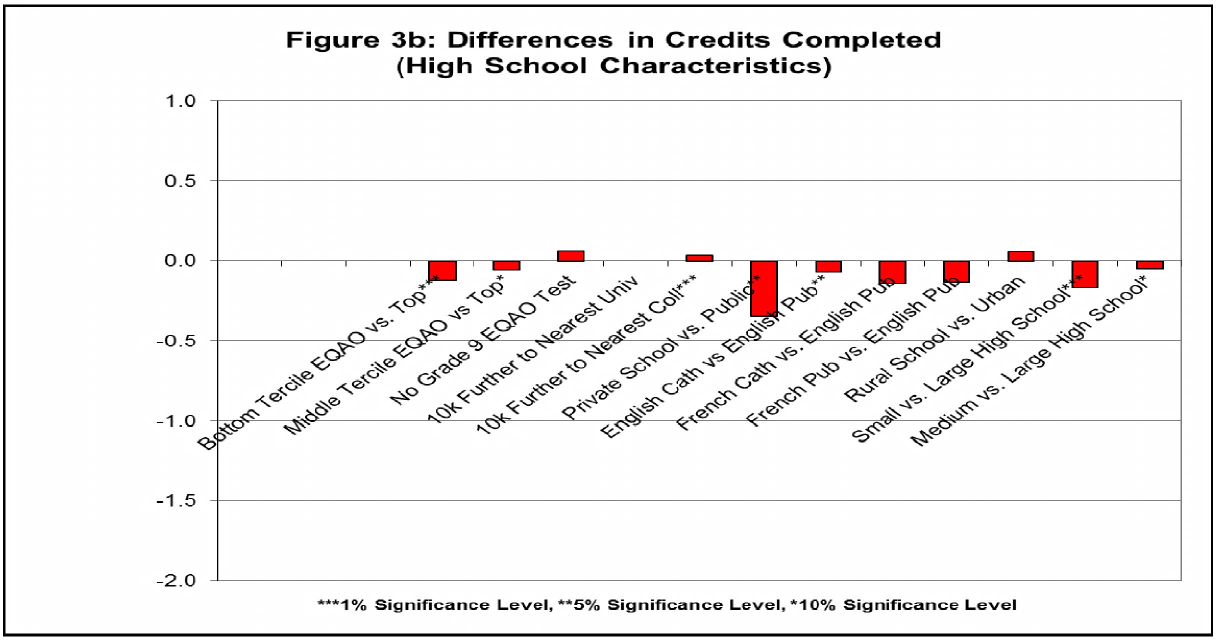


Figure 3 charts the coefficients from the second regression in Table 4 in which the dependent variable is the cumulative credits passed two years after entry. The pattern of coefficients is quite similar to that in Figure 2. Once again, the individual characteristic that stands out most is high school grade average. At the end of two years, students in the lowest high school grade category have earned 2.4 fewer credits (almost a half a year less) than those in the highest high school grade category. The most noticeable difference among programs is that Engineering students have earned almost 0.5 to 0.8 of a credit more than other students. Females have 0.4 credits more than males. The differences in credits passed associated with the remaining individual characteristics are quite small even when statistically significant. One difference from Figure 2 is that English mother tongue is not associated with more credits passed in Figure 3. The estimated effects of high school characteristics in Figure 3b and of neighbourhood characteristics in Figure 3c are very small. For example, the differences between students from high schools in the bottom and top terciles of EQAO scores is only about one-tenth of a credit which is also the difference between students from low- and high-income neighbourhoods.

Figure 3: Difference in Credits





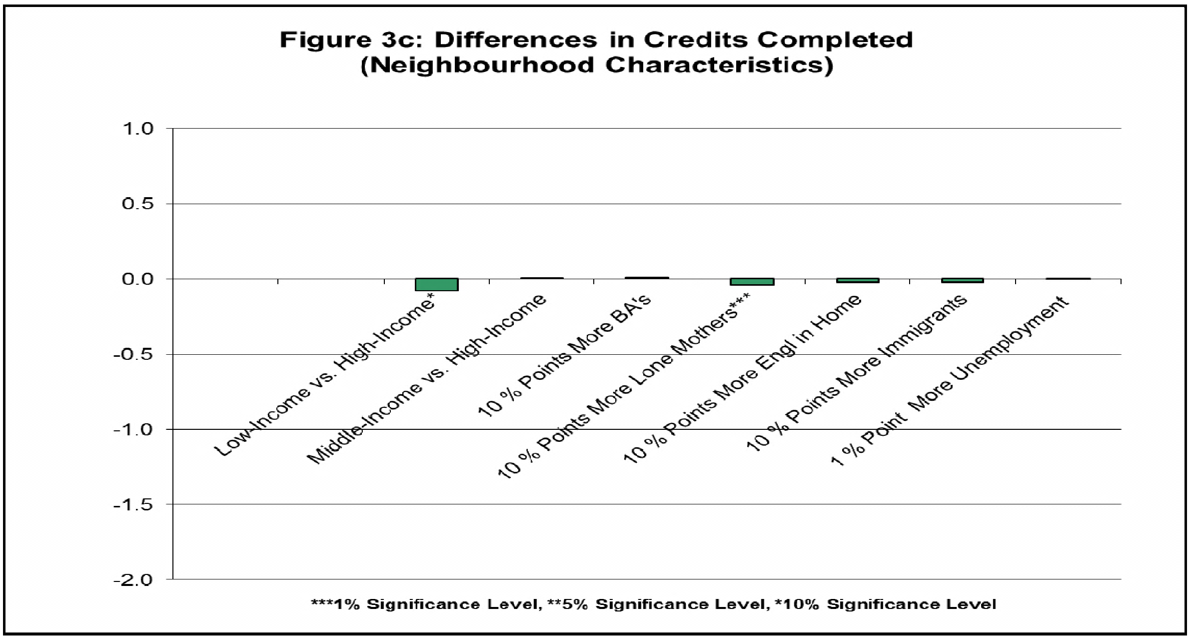
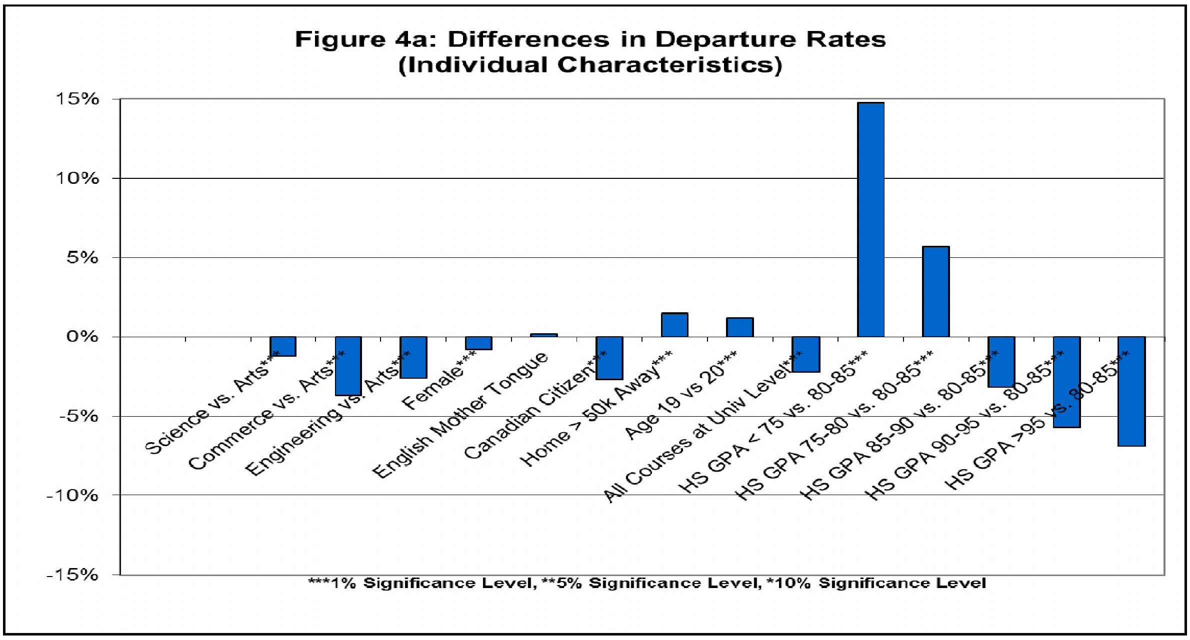
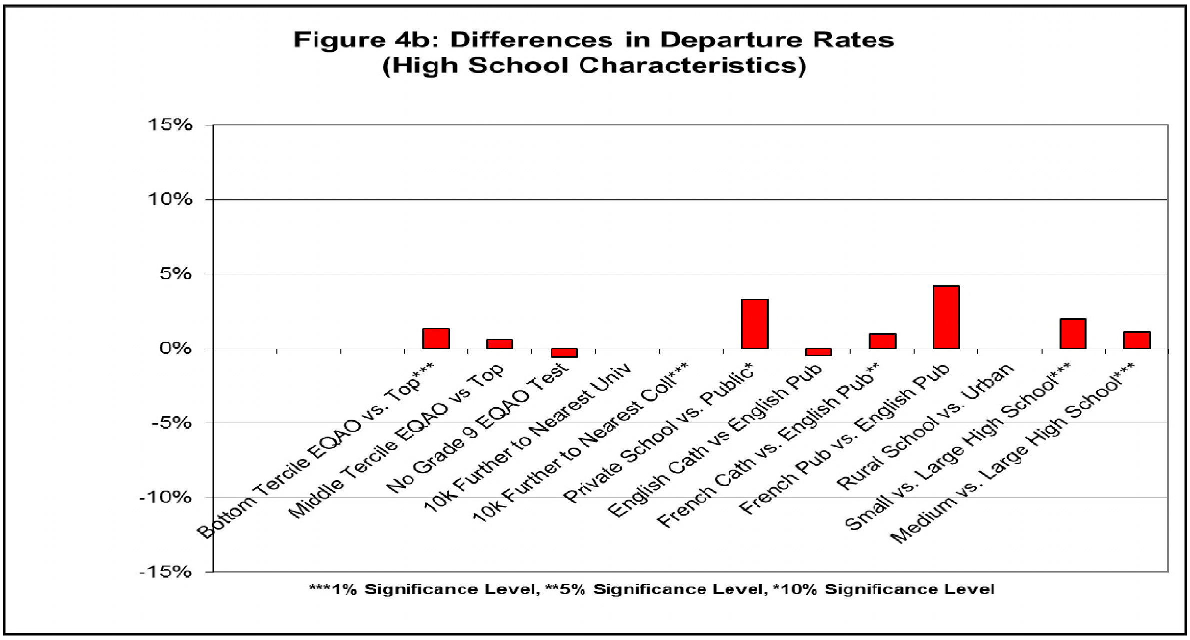


Figure 4 contains the coefficients from the third regression in Table 4 in which the dependent variable is a departure in years one or two after entry (no sign of the student in year 3). The signs of the coefficients are generally the opposite of those in Figures 2 and 3 which is consistent with the interpretation that a departure is generally a “bad” academic outcome rather than a “good” one. The basic message of Figure 4, however, is quite similar to that of the previous two tables. Most importantly, students in the lowest high school grade category have a departure rate that is almost 21 percentage points higher than those in the highest high school grade category. Commerce students have the lowest departure rate which is 4 percentage points lower than that of students in Arts. The female departure rate is less than 1 percentage point lower than that of males. Individual characteristics such as Canadian citizenship and taking all university courses are associated with departure rates that are about 2 percentage points lower, all other factors equal. In Figure 4b, private schools and Francophone public high schools are associated with substantially higher departure rates (about 4 percentage points) but Catholic and rural schools are not. In Figure 4c, neighbourhood characteristics have small impacts. Students from low-income neighbourhoods have a departure rate that is only 1 percentage point higher than that of other students.

Figure 4: Difference in Departures





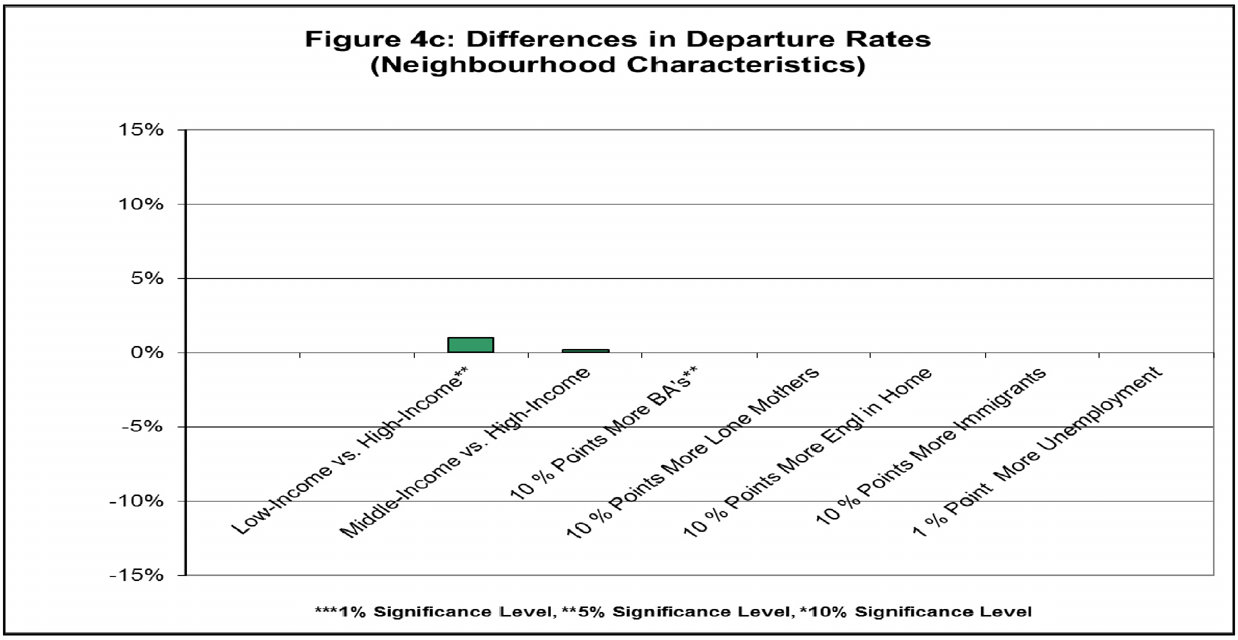
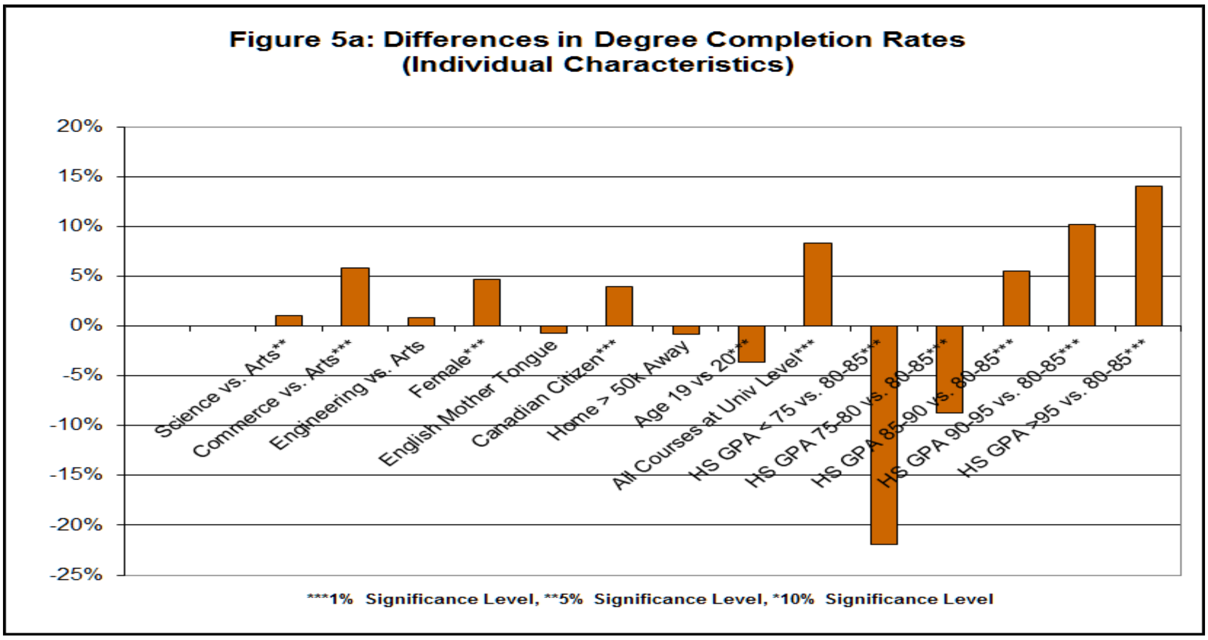
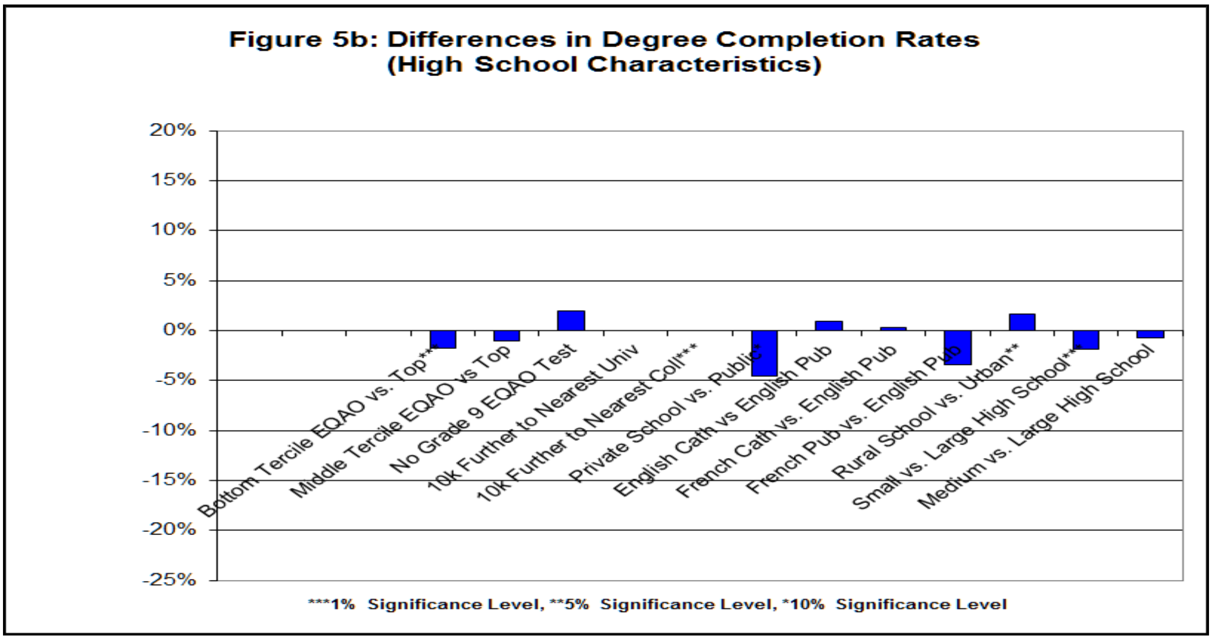


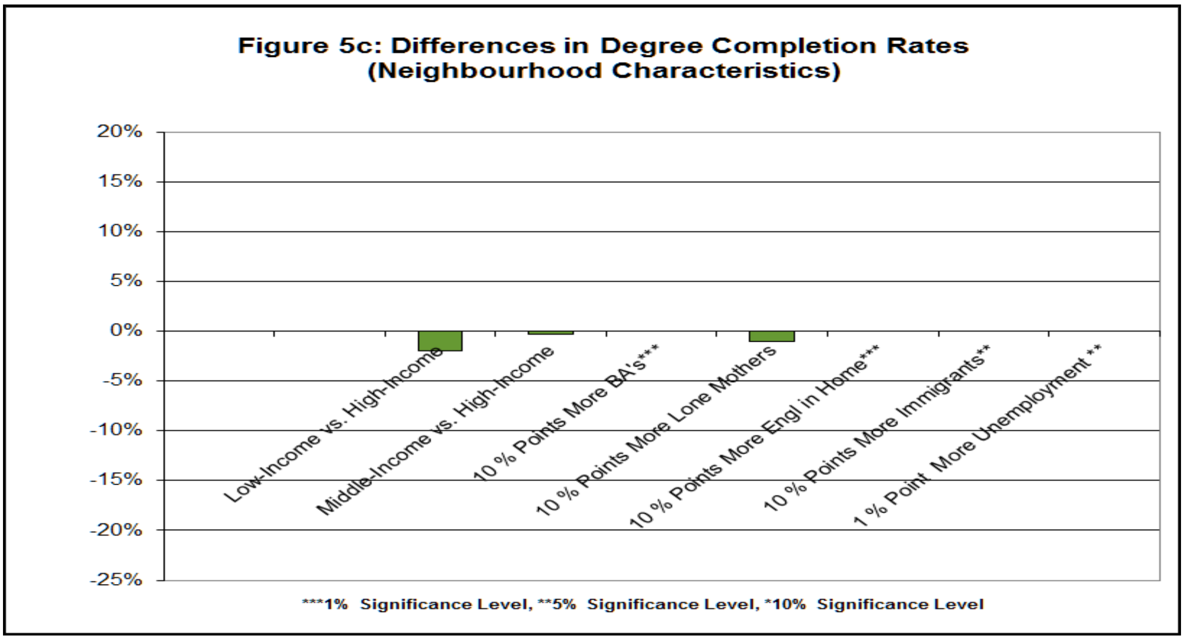
Figure 5 charts the coefficients from the fourth regression in Table 4 in which the dependent variable is degree completion within six years after entry. The qualitative pattern of coefficients is quite similar to that in Figures 2 and 3 (and the reverse of those in Figure 4). Students in the lowest high school grade category have a degree completion rate that is 36 percentage points lower than those in the highest high school grade category. Commerce students have a degree completion rate that is 5 to 6 percentage points higher than that of students in other programs. The female completion rate is almost 5 percentage points higher than that of males. Canadian citizenship is associated with an increase in the likelihood of degree completion of 4 percentage points.[20] Taking all university courses predicts a degree completion rate that is between 8 and 9 percentage points higher. In Figure 5b, students from private high schools have lower completion rates (almost 5 percentage points) and students from rural high schools have slightly higher completion rate (about 2 percentage points). Other high school characteristics also have small impacts as do the neighbourhood characteristics in Figure 5c. For example, the difference in completion rates between students from high schools in the bottom and top terciles of EQAO scores is only about 2 percentage points which is also the difference between students from low- and high-income neighbourhoods.

[20] This impact of citizenship on departures and degree completion may in part reflect a difference between first- and second- generation immigrants as found by Finnie, Childs and Qiu with the Youth In Transition Survey data (2010).

Figure 5: Difference in Degrees Earned







Each of our regressions contained a dummy for each of our four universities (results not reported in Table 4 and 5) and for each entry year. The coefficient estimates for these variables yield conclusions very similar to those which we drew from the trends in mean outcomes in Figure 1 in the following two ways. First, the differences among our four universities are small and their rank order changes. Second, the levels of our persistence measures have generally been quite stable. The one difference from the trends in Figure 1 is that the year dummy coefficients in Tables 4 and 5 only indicate a minor improvement in departure rates for the “double cohort” that entered in 2003. These differences are likely due to the fact that the regressions control for the high school grade averages of entering students.

The regressions in Table 5 contain interactions between high school grades and the following variables: gender, program, neighbourhood income and high school average EQAO performance. We illustrate these interaction effects in Figures 6 through 9. The coefficients for variables that were not interacted with high school grades are very similar in Tables 4 and 5. Figures 6a through 6d present the relationship between high school grade average and each of our outcomes for both males and females. In each figure, we started with the sample mean or proportion of the dependent variable in our sample for males with a high school grade average equal from 80 to 85 and then drew the figures using the estimated coefficients from Table 5 to calculate the estimated values for the other high school grade averages.[21]

[21] We centered the figures at the 80 to 85 grade range because this contains the sample mean (83.4).

	Table 5 Regressions With Interactions				
	(1)	(2)	(3)	(4)	(5)
	Dependent Variable	Cumulative Grade Average After Two Years	Credits Passed After Two Years	Departed During Years 1 & 2	Completed Degree Within 6 Years
(1)	Science Entry Program	-1.051***	0.183***	-0.021***	0.027***
		(0.125)	(0.031)	(0.005)	(0.007)
(2)	Business Entry Program	-0.623***	0.322***	-0.044***	0.069***
		(0.143)	(0.047)	(0.006)	(0.009)
(3)	Engineering Entry Program	-3.711***	1.035***	-0.015*	0.000
		(0.195)	(0.064)	(0.008)	(0.013)
(4)	Female	1.111***	0.450***	-0.016***	0.055***
		(0.092)	(0.029)	(0.004)	(0.007)
(5)	English Mother Tongue	0.646***	0.001	0.002	-0.006
		(0.122)	(0.030)	(0.004)	(0.006)
(6)	Canadian Citizen	0.722***	0.181**	-0.027***	0.040***
		(0.272)	(0.074)	(0.007)	(0.009)
(7)	Home is 50 km Or More From University	-0.800***	-0.184***	0.016***	-0.008
		(0.166)	(0.029)	(0.004)	(0.006)
(8)	Age at Entry (months)	-0.078***	-0.017***	0.001***	-0.003***
		(0.007)	(0.002)	(0.000)	(0.000)
(9)	Best Six Grade 12/13 Courses All University	1.073***	0.197***	-0.021***	0.089***
		(0.114)	(0.033)	(0.005)	(0.021)
(10)	HS Average Grade < 75	-8.038***	-1.685***	0.143***	-0.217***
		(0.232)	(0.083)	(0.013)	(0.017)
(11)	HS Average Grade =>75 and <80	-4.199***	-0.644***	0.044***	-0.086***
		(0.163)	(0.053)	(0.008)	(0.013)
(12)	HS Average Grade =>85 and <90	5.166***	0.405***	-0.031***	0.059***
		(0.178)	(0.051)	(0.007)	(0.013)
(13)	HS Average Grade =>90 and <95	11.214***	0.864***	-0.046***	0.124***
		(0.235)	(0.061)	(0.009)	(0.013)
(14)	HS Average Grade =>95	17.696***	1.150***	-0.047**	0.168***
		(0.436)	(0.110)	(0.021)	(0.027)
(15)	Low Income EA/DA	-0.170	-0.124**	0.015**	-0.029***
		(0.198)	(0.054)	(0.006)	(0.009)
(16)	Middle Income EA/DA	-0.029	-0.019	0.003	-0.010
		(0.121)	(0.033)	(0.005)	(0.008)
(17)	% EA/DA Bachelor's Degree	0.018***	0.001	-0.000**	0.000**
		(0.006)	(0.001)	(0.000)	(0.000)
(18)	% EA/DA Lone Mother Families	-0.006	-0.004***	0.000*	-0.001***
		(0.005)	(0.001)	(0.000)	(0.000)
(19)	% EA/DA English Mother Tongue	-0.019***	-0.002*	-0.000	0.000
		(0.006)	(0.001)	(0.000)	(0.000)
(20)	% EA/DA Immigrant since 1981	-0.013**	-0.002	-0.000	0.000
		(0.006)	(0.001)	(0.000)	(0.000)
(21)	% EA/DA Unemployed	-0.017*	-0.003	0.000	0.000
(22)	% High Scores on Grade 9 EQAO Test in Bottom Tercile	-0.932***	-0.151***	0.020***	-0.018**
		(0.225)	(0.046)	(0.006)	(0.008)
(23)	% High Scores on Grade 9 EQAO Test in Middle Tercile	-0.656***	-0.068*	0.004	-0.009
		(0.209)	(0.041)	(0.005)	(0.008)

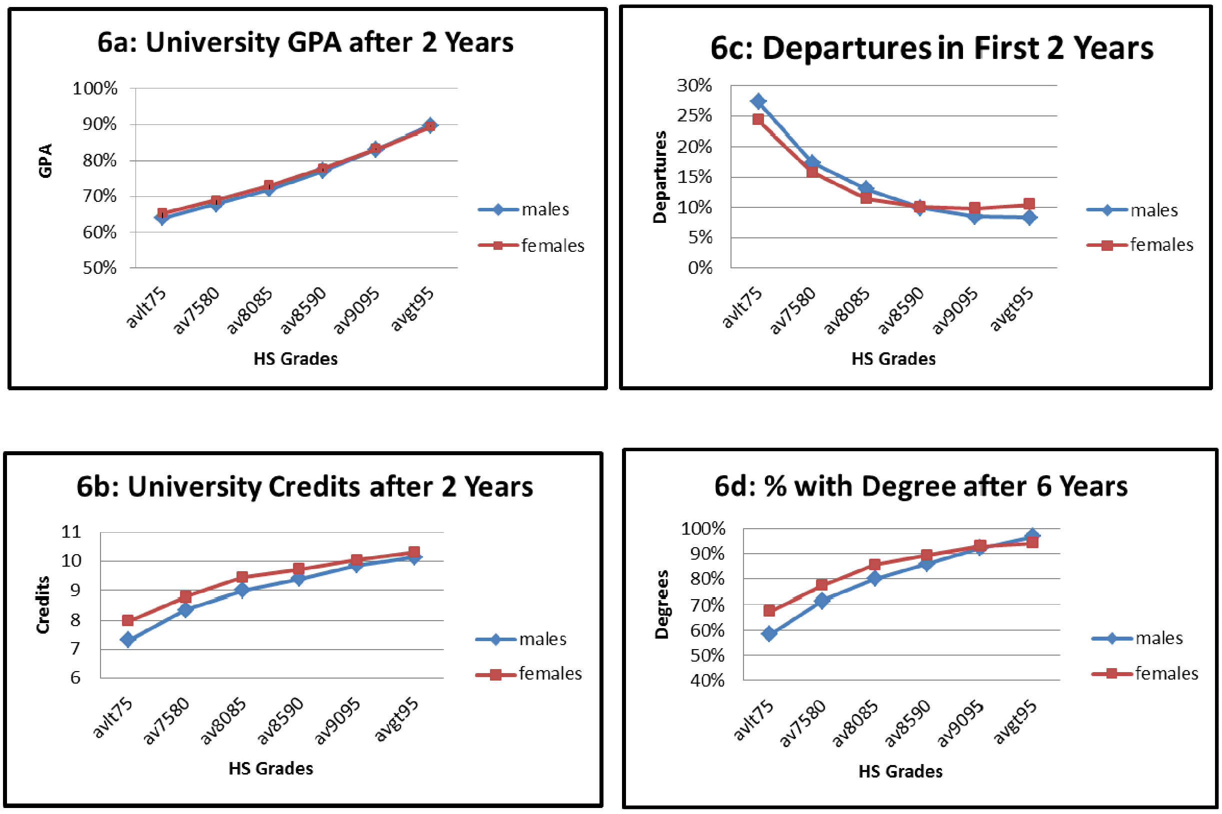
(23)	% High Scores on Grade 9 EQAO Test in Middle Tercile	-0.656***	-0.068*	0.004	-0.009
		(0.209)	(0.041)	(0.005)	(0.008)
(24)	No Grade 9 EQAO Test	1.705***	0.073	-0.010	0.023
		(0.625)	(0.127)	(0.018)	(0.023)
		Table 5 (continued)			
(25)	Distance of High School from Nearest University (km)	-0.002	0.000	-0.000	0.000
		(0.002)	(0.000)	(0.000)	(0.000)
(26)	Distance of High School from Nearest College (km)	0.020***	0.003***	-0.000***	0.000***
		(0.004)	(0.001)	(0.000)	(0.000)
(27)	Private High School	-3.078***	-0.376**	0.037*	-0.049*
		(0.765)	(0.160)	(0.019)	(0.025)
(28)	English Catholic High School	-1.023***	-0.075**	-0.005	0.009
		(0.178)	(0.033)	(0.004)	(0.006)
(29)	Francophone Public High School	1.237***	-0.144	0.043**	-0.036
		(0.472)	(0.159)	(0.019)	(0.034)
(30)	Francophone Catholic High School	-0.365	-0.152	0.012	0.001
		(0.628)	(0.132)	(0.014)	(0.023)
(31)	Rural High School	0.357*	0.048	0.000	0.017**
		(0.211)	(0.041)	(0.006)	(0.007)
(32)	High School Enrolment in Bottom Tercile	-0.914***	-0.144***	0.018***	-0.018**
		(0.242)	(0.045)	(0.005)	(0.007)
(33)	High School Enrolment in Middle Tercile	-0.398**	-0.042	0.010**	-0.007
		(0.194)	(0.032)	(0.004)	(0.006)
(34)	1994 Entry Year	0.000	0.170***	-0.004	0.000
		(0.000)	(0.059)	(0.009)	(0.000)
(35)	1995 Entry Year	-0.359***	0.095*	0.012	-0.019***
		(0.134)	(0.057)	(0.009)	(0.007)
(36)	1996 Entry Year	-0.219	0.149**	0.002	-0.009
		(0.144)	(0.060)	(0.009)	(0.007)
(37)	1997 Entry Year	0.048	0.198***	-0.009	-0.004
		(0.143)	(0.057)	(0.009)	(0.008)
(38)	1998 Entry Year	0.231	0.240***	-0.006	-0.004
		(0.143)	(0.052)	(0.009)	(0.007)
(39)	1999 Entry Year	0.264*	0.119**	-0.005	-0.004
		(0.137)	(0.050)	(0.008)	(0.007)
(40)	2000 Entry Year	0.329**	0.199***	-0.013*	-0.006
		(0.151)	(0.051)	(0.008)	(0.009)
(41)	2001 Entry Year	0.450***	0.220***	-0.011	-0.018*
		(0.156)	(0.051)	(0.007)	(0.011)
(42)	2002 Entry Year	0.403***	0.222***	-0.010	0.000
		(0.151)	(0.047)	(0.008)	(0.000)
(43)	2003 Entry Year	0.241	0.094**	-0.022***	0.000
		(0.161)	(0.045)	(0.007)	(0.000)
(44)	2004 Entry Year	-0.773***	-0.046	0.000	0.000
		(0.213)	(0.043)	(0.000)	(0.000)
(45)	2005 Entry Year	-0.806***	0.000	0.000	0.000
		(0.233)	(0.000)	(0.000)	(0.000)
(46)	Female*Average Grade Less Than 75	0.235	0.212***	-0.013	0.036**
		(0.186)	(0.066)	(0.010)	(0.015)

(47)	Female*Average Grade 75-80	0.036	-0.008	0.000	0.006
		(0.136)	(0.046)	(0.007)	(0.011)
(48)	Female*Average Grade 85-90	-0.436***	-0.113***	0.019***	-0.021**
		(0.135)	(0.037)	(0.006)	(0.009)
(49)	Female*Average Grade 90-95	-1.006***	-0.267***	0.029***	-0.049***
		(0.161)	(0.043)	(0.006)	(0.009)
(50)	Female*Average Grade 95 or More	-1.565***	-0.281***	0.038***	-0.082***
		(0.264)	(0.075)	(0.009)	(0.014)
(51)	Science*Average Grade Less Than 75	0.670***	0.070	0.051***	-0.041**
		(0.230)	(0.086)	(0.014)	(0.017)
(52)	Science*Average Grade 75-80	0.419**	0.117**	0.024***	-0.026**
		(0.162)	(0.052)	(0.008)	(0.012)
(53)	Science*Average Grade 85-90	-0.457***	-0.065	-0.010	0.000
		(0.131)	(0.042)	(0.007)	(0.011)
(54)	Science*Average Grade 90-95	-0.051	-0.174***	-0.008	-0.028**
		(0.176)	(0.051)	(0.008)	(0.012)
(55)	Science*Average Grade 95 or More	1.129***	-0.384***	-0.016	-0.034
		(0.424)	(0.102)	(0.021)	(0.026)
(56)	Business*Average Grade Less Than 75	0.261	0.260	-0.011	0.017
		(0.513)	(0.211)	(0.028)	(0.036)
(57)	Business*Average Grade 75-80	-0.010	-0.018	0.009	-0.004
		(0.220)	(0.072)	(0.011)	(0.016)
(58)	Business*Average Grade 85-90	-0.667***	-0.072	-0.003	-0.002
		(0.168)	(0.050)	(0.008)	(0.013)
(59)	Business*Average Grade 90-95	-0.580**	-0.302***	0.010	-0.040***
		(0.244)	(0.072)	(0.010)	(0.015)
(60)	Business*Average Grade 95 or More	0.778	-0.531***	-0.027	-0.022
		(0.583)	(0.122)	(0.022)	(0.027)
(61)	Engineering*Average Grade Less Than 75	1.686**	-0.257	-0.036	-0.037
		(0.726)	(0.335)	(0.036)	(0.051)
(62)	Engineering*Average Grade 75-80	0.571*	-0.402***	0.048***	-0.074***
		(0.314)	(0.125)	(0.017)	(0.027)
(63)	Engineering*Average Grade 85-90	-0.666***	-0.086	-0.028***	0.017
		(0.230)	(0.081)	(0.010)	(0.017)
(64)	Engineering*Average Grade 90-95	-1.072***	-0.658***	-0.041***	0.036**
		(0.272)	(0.085)	(0.011)	(0.017)
(65)	Engineering*Average Grade 95 or More	-1.164**	-0.865***	-0.053**	0.028
		(0.477)	(0.124)	(0.021)	(0.026)
(66)	Low Income*Average Grade Less Than 75	0.619**	0.041	-0.015	0.006
		(0.280)	(0.096)	(0.014)	(0.020)
(67)	Low Income*Average Grade 75-80	0.123	-0.001	0.001	0.010
		(0.190)	(0.070)	(0.010)	(0.015)
(68)	Low Income*Average Grade 85-90	-0.133	0.058	-0.000	0.013
		(0.167)	(0.051)	(0.007)	(0.012)
(69)	Low Income*Average Grade 90-95	-0.226	0.190***	-0.022***	0.016
		(0.211)	(0.061)	(0.008)	(0.013)
(70)	Low Income*Average Grade 95 or More	-1.080***	0.088	-0.021*	0.058***
		(0.403)	(0.104)	(0.013)	(0.018)
(71)	Middle Income*Average Grade Less Than 75	0.568***	0.031	-0.006	0.011
		(0.204)	(0.077)	(0.012)	(0.018)

(72)	Middle Income*Average Grade 75-80	0.422***	0.081	-0.002	0.010
		(0.131)	(0.050)	(0.008)	(0.012)
(73)	Middle Income*Average Grade 85-90	-0.308**	-0.001	0.004	0.010
		(0.133)	(0.040)	(0.006)	(0.010)
(74)	Middle Income*Average Grade 90-95	-0.250	0.022	-0.009	0.017
		(0.163)	(0.047)	(0.007)	(0.011)
(75)	Middle Income*Average Grade 95 or More	-0.208	-0.028	-0.011	0.014
		(0.302)	(0.079)	(0.010)	(0.015)
(76)	Bottom Tercile EQAO*Average Grade Less Than 75	-0.285	-0.171	0.006	-0.036*
		(0.294)	(0.117)	(0.015)	(0.020)
(77)	Bottom Tercile EQAO*Average Grade 75-80	0.218	-0.063	-0.004	0.007
		(0.189)	(0.075)	(0.011)	(0.014)
(78)	Bottom Tercile EQAO*Average Grade 85-90	-0.206	0.150***	-0.015**	0.001
		(0.162)	(0.052)	(0.007)	(0.011)
(79)	Bottom Tercile EQAO*Average Grade 90-95	-0.682***	0.073	-0.025***	0.010
		(0.221)	(0.060)	(0.007)	(0.012)
(80)	Bottom Tercile EQAO*Average Grade 95 or More	-0.505	0.186*	-0.020*	0.002
		(0.441)	(0.095)	(0.012)	(0.019)
(81)	Middle Tercile EQAO*Average Grade Less Than 75	-0.233	-0.162*	0.019	-0.020
		(0.226)	(0.089)	(0.013)	(0.017)
(82)	Middle Tercile EQAO*Average Grade 75-80	0.107	-0.066	0.011	0.007
		(0.153)	(0.053)	(0.008)	(0.012)
(83)	Middle Tercile EQAO*Average Grade 85-90	0.042	0.054	0.000	-0.005
		(0.144)	(0.044)	(0.006)	(0.009)
(84)	Middle Tercile EQAO*Average Grade 90-95	-0.022	0.085	-0.014**	0.007
		(0.220)	(0.057)	(0.007)	(0.011)
(85)	Middle Tercile EQAO*Average Grade 95 or More	0.460	0.260***	-0.006	0.006
		(0.352)	(0.086)	(0.011)	(0.016)
(86)	Constant	89.353***	13.068***	-0.106**	1.300***
		(1.487)	(0.414)	(0.048)	(0.081)
	Observations	113,271	113,407	97,558	55,574
	R-squared	0.382	0.115	0.044	0.075
	Sample mean or proportion	72	9.2	13%	80%
	Clustered standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1				
	Constant (reference group) is for a male, Arts, non-English mother tongue, not a citizen, resides within 50 km, some non-University courses, GPA 80-85, high income, top tercile EQAO scores, 2006 entry. High school is publicly funded, English, public and urban. Each regression also contains a dummy variable for each university.				

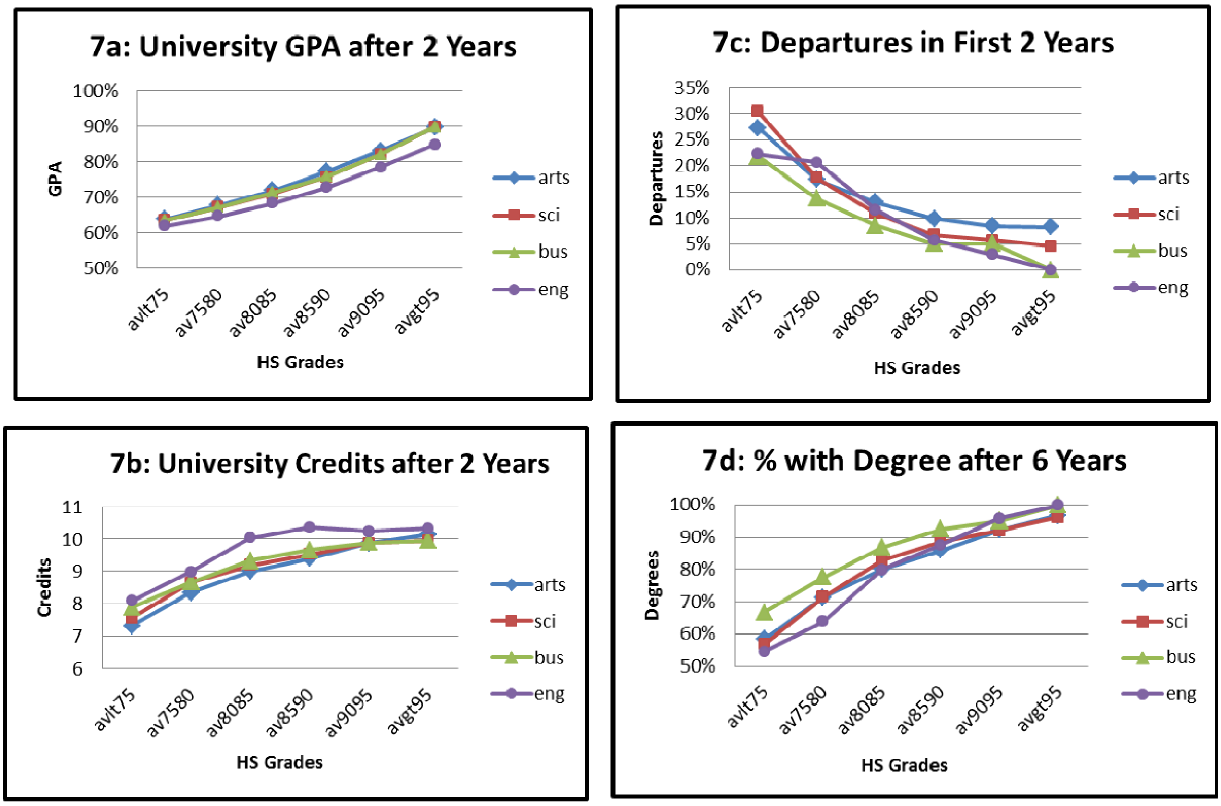
Figures 6a through 6d indicate that the strong relationship between high school grade category and university outcomes is true of both males and females. These figures also reveal an interesting pattern of gender differences. Among students with low high school grades, females have better university outcomes for most of the measures. For example, among students with a high school average of less than 75, the female departure rate is 3 percentage points lower than that of males and the female degree completion rate is 9 percentage points higher. The performance levels of females and males converge or even reverse as one moves to the right on the charts to better high school grade averages. Among students with a high school average of 95 or better, for example, the female departure rate is 2.5 percentage points higher than that of males and the female degree completion rate is 3 percentage points lower (both showing worse female outcomes). Although less easy to discern from the chart, the grade point averages for males and females follow a similar pattern with a female advantage of about 1.4 percentage points for students with the lowest high school averages and a negligible difference at the top end of the high school grade distribution.

Figure 6: Persistence Measures by Gender and High School Average



Figures 7a through 7d show the relationship between high school grade averages and university outcomes by entry program (or faculty). Figures 7a and 7b shows that Engineering students have lower grade averages and more credits completed after two years than students in other programs regardless of high school grade category. Figure 7c indicates that Business students have the lowest departure rates in most grade categories. Engineering students also have low departure rates for high school grade averages of less than 75 and above 85. Our regression even predicts very small negative departure rates for Business and Engineering students with a high school grade of 95 and over though in the graph we force the minimums to be no lower than zero.[22] Figure 7d indicates that Business students have the highest degree completion rates in most grade categories. Engineering students have lowest degree completion rates for high school grade averages under 85 but high degree completion rates for high school grade averages above 90.

Figure 7: Persistence Measures by Program and High School Average

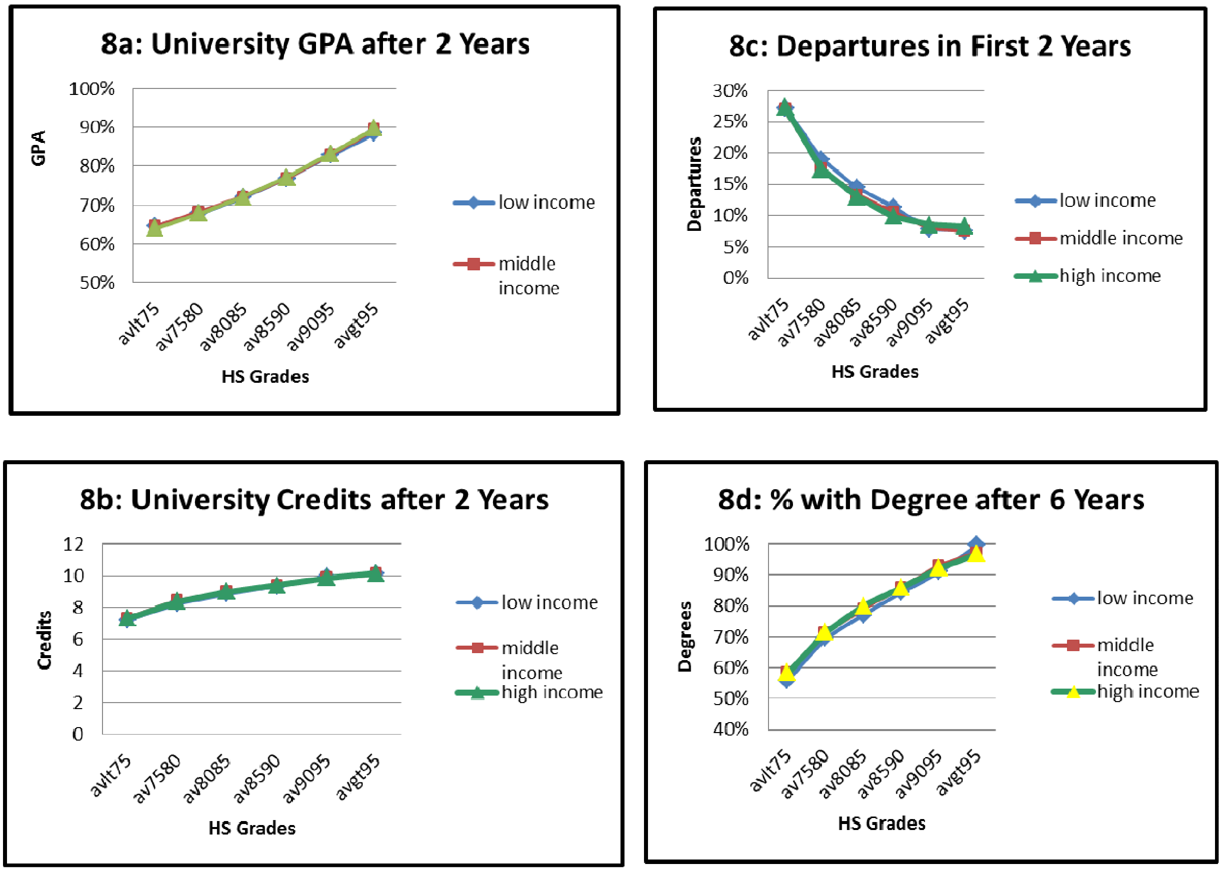


Figures 8a through 8dshow the relationship between high school grade averages and university outcomes by neighbourhood average equivalent income. In Figures 8a and 8b, the differences between students from low income and high income neighbourhoods in university grades are 1 percentage point or less and the differences in credits completed are 0.2 credits or less for each high school grade category. Departure rates in Figure 8c show slightly larger differences and an interesting pattern. Students with low high school grades have slightly higher departure rates than high income students and the reverse is true among students with a high school grade average of 90 or more. All such differences, however, are at most 1.5 percentage points. The differences in degree completion rates in Figure 8d between low- and high-income students are at most 3 percentage points and tend to be larger at lower high school grade levels. These

[22] A negative departure rate is, of course, not possible but this prediction does reflect very low departure rates in reality for such students. In the same vein, the next regression predicts a degree completion rate greater than 100% for Business and Engineering students with a high school average of 95 or higher.

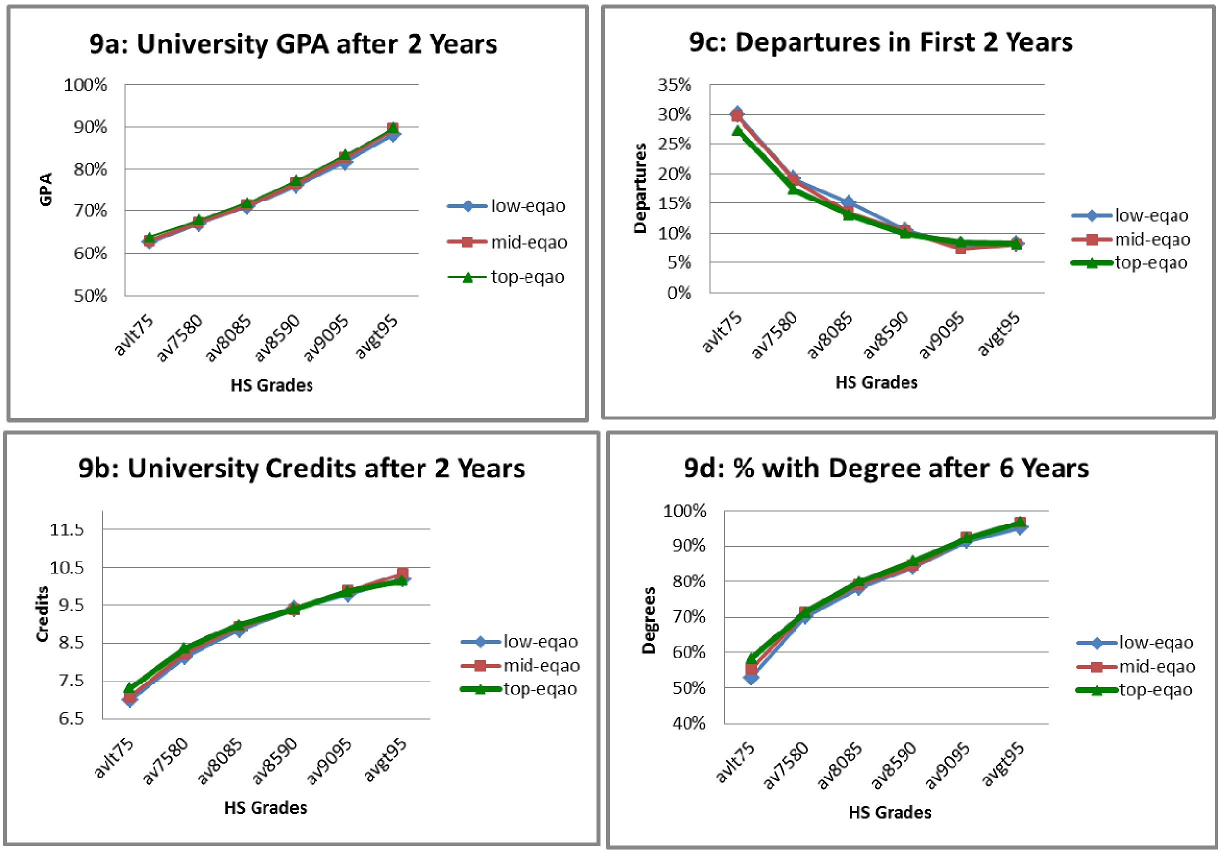
last two measures suggest that some lower income students may be forced to ‘go slower’, perhaps dropping out for financial reasons and thence graduating more slowly, if at all. Figure 8 implies that neighbourhood income does not have a substantial link with our outcomes regardless of high school grade average. As indicated above, Dooley, Payne and Robb (2009) found that average neighbourhood income has a much stronger link with applications to university. This distinction provides an important insight for future research and we return to this optic in the final section of the paper.

Figure 8: Persistence Measures by Income Tercile and High School Average



Finally, Figures 9a through 9d show the relationship between high school grade averages and university outcomes by the average performance of the student’s high school on the academic Grade 9 EQAO test. As with neighbourhood average equivalent income, the differences by high school performance are slight. In Figures 9a and 9b, the differences in university grade averages between individuals from schools with the smallest proportion of high EQAO scores and those from schools with the largest proportion of high EQAO scores are all 1.5 percentage points or less and the differences in credits completed are all 0.3 credits or less. The same differences in departure rates in Figure 9c are at most 2 percentage points and the differences in degree completion rates in Figure 9d are at most 3 percentage points. As with differences by neighbourhood income, the difference by high school EQAO performance are most pronounced among students with lower high school grade averages.

Figure 9: Persistence Measures by EQAO Tercile and High School Average



We have estimated a variety of regression model specifications other than those in Tables 4 and 5. In one alternative specification, we included a fixed effect (dummy variable) for each high school in our sample and, hence, excluded the variables measuring high school characteristics. The inclusion of the high school fixed effects increases the R squared, as expected, but does not have a noticeable effect on the coefficient estimates for individual or neighbourhood characteristics. In a second alternative specification, we omitted high school grades from the regression. As expected, this omission decreases the R squared greatly, especially in the regression for university grade averages, but does not have a noticeable effect on the coefficient estimates for individual, neighbourhood or high school characteristics. The estimates of these alternative specifications are available upon request.

Summary and Conclusions

In this paper, we address a series of questions concerning academic persistence and success among university students. We use administrative data that have been collected on students in four Ontario universities and have linked these data with information on the students' individual characteristics (including high school performance), neighbourhood, and high school. The students are members of cohorts that came directly from Ontario high schools and entered one of our four universities for full-time degree study in the fall of 1994 through 2006.

We considered a wide variety of measures of persistence but focused the analysis in this paper on the following four: cumulative grade average and credits completed at the end of Year 2 after university entry, departures during Year 1 and Year 2 after entry, and degrees completed within six years after entry. This rich data source clearly merits more extensive analysis but we believe that a number of important lessons can be inferred from the results reported thus far.

First, the time trends reveal that the values of all four outcomes have generally been quite stable over time at each of our institutions. This stability over time in the levels of our measures of academic success in university is true of both the simple means of the variables and when we control statistically for a wide variety of individual, neighbourhood and school characteristics.

Second, academic performance (grade point average) in secondary school is strongly linked to all of our measures of university performance even those, such as degree attainment, that occur years after the students have received their high school diploma. These links are strong in the sense of both the magnitude and the precision of the estimated coefficients. Furthermore, the explanatory power of the high school grade average greatly dominates that of other variables such as university program, gender, and neighbourhood and high school characteristics.

Third, the neighbourhood and high school characteristics used in this study, such as average income and EQAO scores, have relatively weak links with our measures of persistence. In contrast, Dooley, Payne and Robb (2009) reported that such measures have much stronger links with the decision to apply to university. Students from low income neighbourhoods are 14 percentage points less likely to apply to university than those in high income neighbourhoods (controlling for other factors including GPA). Students from high schools in which less than 50% of students attained a high EQAO score (3 or 4) were 8 percentage points less likely to apply to university than students in high schools in which more than 50% received a high EQAO score (controlling for other factors including GPA). Hence, it appears that neighbourhood and high school characteristics play a larger role in the application decision than in the achievement of academic success once attending university. We hasten to add, however, several important qualifications to this conclusion: (a) numerous small differences in persistence can still have a large cumulative impact on the life of a young person; (b) our current measures of neighbourhood and high school characteristics are quite limited; and (c) variation in the socio-economic characteristics of individual families within neighbourhoods and in the academic characteristics of academic programs within high schools may have substantial success in accounting for variation in university persistence.

Fourth, our results point to the advantages of rich administrative data. Such data suffer much less than do survey data from response and selection bias. We do not have the students' reports of their grades but their actual grades and we have close to full coverage of the relevant populations rather than just responses to a voluntary survey. In addition, the number of observations is usually much larger in administrative data sets than in surveys.

We currently have a detailed picture of the individual student while in university and these data have much more to tell us. For example, one could study additional outcome measures, such as graduation average, and analyze all outcome measures by university program, academic level, gender, mother tongue, type of high school, etc. Our data challenge is the very limited information that we currently have concerning the individual student prior to university. As stressed above, our current set of neighbourhood and high school characteristics are limited as is our understanding of what lies behind the large estimated impact of high school grades. Many additional insights could be gained by linking our current data with other administrative data sets. Individual information concerning income and other family characteristics is potentially available via linkage with OSAP records. Individual student records from the Ministry of Education could shed light on the decisions and strategies in high school that are associated with both a strong high school grade average and success in university.

References

Bowlby, Jeffrey and Kathryn McMullen. 2002. "At a Crossroads: First Results for the 18 to 20-Year-old Cohort of the Youth in Transition Survey." Ottawa: Human Resources Development Canada and Statistics Canada.

Chemin, Matthieu. 2009. "Does student financial aid cause more participation in and graduation from post-secondary education?" MESA Project Research Paper. Toronto, ON: Educational Policy Institute.

Conrad, Mark and Katherine Morris 2010. "Shifting from Retention Rates to Retention Risk: An Alternative Approach for Managing Insitutional Student Retention Performance". Toronto: Higher Education Quality Council of Ontario.

Dooley, Martin, A. Abigail Payne and A. Leslie Robb. 2009. "University Participation and Income Differences: An Analysis of Applications by Ontario Secondary School Students." Toronto, ON: Higher Education Quality Council of Ontario.

Finnie, Ross and Theresa Qiu. 2008. "Is The Glass (or Classroom) Half-Empty or Nearly Full? New Evidence on Persistence in Post-Secondary Education in Canada." In Who Goes? Who Stays? What Matters? Accessing and Persisting in Post-Secondary Education in Canada eds. R. Finnie, R.E. Mueller, A. Sweetman and A. Usher. Montreal and Kingston: McGill-Queens University Press.

Finnie, Ross and Theresa Qiu. 2009. "Moving Through, Moving On: Persistence in Post-Secondary Education in Atlantic Canada, Evidence from the PSIS." Ottawa: Statistics Canada, Centre for Education Statistics Research Papers.

Finnie, Ross, Stephen Childs and Theresa Qiu. 2010. "Patterns of Persistence in Postsecondary

Education: New Evidence for Ontario." The Education Policy Research Initiative, University of Ottawa.

Grayson, J. Paul and Kyle Grayson. 2003. "Research on Retention and Attrition." Montreal: Canadian Millenium Scholarship Foundation.

Johnson, David. 2008. "How is Variation in Tuition Across Canadian Provinces Related to University Participation in the Youth in Transition Survey?" In Who Goes? Who Stays? What Matters? Accessing and Persisting in Post-Secondary Education in Canada eds. R. Finnie, R.E. Mueller, A. Sweetman and A. Usher. Montreal and Kingston: McGill-Queens University Press.

Martinello, Felice. 2008. "Transitions and Adjustments in Students' Canadian Post-Secondary Education." In Who Goes? Who Stays? What Matters? Accessing and Persisting in

Post-Secondary Education in Canada eds. R. Finnie, R.E. Mueller, A. Sweetman and A. Usher. Montreal and Kingston: McGill-Queens University Press.

Mueller, Richard E. 2008a. "Access and Persistence of Students in Canadian Post-Secondary Education: What We Know, What We Don't Know and Why It Matters." In Who Goes? Who Stays? What Matters? Accessing and Persisting in Post-Secondary Education in Canada eds. R. Finnie, R.E. Mueller, A. Sweetman and A. Usher. Montreal and Kingston: McGill-Queens University Press.

Mueller, Richard E. 2008b. "Access and Persistence of Students from Low-Income Backgrounds in Canadian Post-Secondary Education: A Review of the Literature." MESA Project Research Paper 2008-1. Toronto, ON: Educational Policy Institute.

Nory, Amaury, Elizabeth Barlow, and Gloria Crisp. 2005. "Student Persistence and Degree Attainment Beyond the First Year in College: The Need for Research." In College Student Retention: Formula for Student Success. Ed. Alan Seidman. Westport, CT: Praeger Publishers

Ontario Ministry of Training, Colleges and Universities 2010. "Key Progress Indicators for Universities."

Parkin, Andrew, and Noel Baldwin 2009. "Persistence in Postsecondary Education in Canada: The Latest Research", Research Note #8. Montreal: Canada Millenium Scholarship Foundation.

Data Appendix

Introduction

The primary purpose of this appendix is to document the building of the data set used for the analysis of persistence and academic success among students at four Ontario universities. This appendix documents key data issues encountered and our resolutions of those issues. We make use of the following four administrative data sources.

- Administrative data from four Ontario universities for varying periods from 1990-2006
  - The observations are at the student level and cover varying time periods
  - The extracts include students graduating from an Ontario high school and starting a full-time degree program in the following Fall semester
  - We distinguish a year in these data as the year in which a given school year starts (e.g. year 1999 stands for school year 1999/2000 which starts in the fall of 1999 and ends in the spring of 2000)

- Ontario University Application Centre (OUAC) data from 1994-2006
  - Each observation represents a student level application for full time admission to any Ontario university from students in high school (called 101 students, which make up the bulk of all applications)
  - A year in this data is measured as the year in which a student would enroll in university (e.g. 1994 application year is for students starting university in September 1994). The application would have been made in the previous fall

- Census Data
  - We use the postal code of the student's home residence at time of the OUAC application to link to data on neighbourhood characteristics from the Enumeration/Dissemination Area for the corresponding Census (1996, 2001 or 2006). These neighbourhood characteristics provide indicators of the socioeconomic background of the student.

- High School Data from 2001 through 2004
  - We use measures obtained from the Ministry of Education to reflect the characteristics of the high schools from which a student graduates, such as school size, location, and performance on standardized tests.
  - 2001 through 2004 are the years for which we currently have complete data.

University Data Requested and Provided

Initial Request Made to the Universities

Requests were originally made to all Ontario universities. Four universities provided us with comparable data.

Students for Whom Information was Requested:

- We requested information on student cohorts for 1993 to the most recently available, where we define a cohort as all students who registered in Level 1 at a particular Ontario University in the Fall of a given calendar year. We requested information for only those students entering a full time university degree program directly from an Ontario high school.   This limitation corresponded to the information available from OUAC and are referred by the name university registrars and OUAC use for them, '101 students'.

Admission (and one time information) requested:

- OUAC application number
- Program in which the student registered and the student's final admission average for that program
- Amount of entry scholarship offered
- Total value of all years of scholarship if any required average is maintained (could be same as above if no renewal is guaranteed)
- Average grade for scholarship purposes if different from admission average

Graduation information requested:

- Degree(s) awarded by date awarded, type of degree and graduating grade point average.

Academic Progress information requested:

For each student, we requested all of the following variables for each semester in which the student was registered as an undergraduate including summer terms

- Academic level
- Academic degree program
- Full-time or part-time status
- Amount of scholarship offered
- Academic credits passed during the semester
- Academic credits failed during the semester
- Academic credits cancelled during the semester where cancelled means dropped after the drop/add period (an official record is kept) but before a grade is mandatory

- Overall grade average for the semester and number of courses/units on which this average is computed
- Cumulative grade average
- Cumulative number of credits passed to date
- Number of credits required for completion of current degree program
- Annual (academic year) dollar value of scholarships held by the student

Limitations in Data Provided by the Universities

The data provided by the four universities were quite similar in most regards save for the following three aspects: the period of data availability; the availability of scholarship information; and the degree of data disaggregation.

University	Entry Years Available	Scholarship Data	Academic Terms
University 1	1994-2004	None available	Fall, Winter, Summer
University 2	1994-2005	Entry and continuing	Fall, Winter, Summer
University 3	1999-2006	Entry and continuing	Fall/Winter, Summer
University 4	1994-2004	None available	Fall, Winter, Summer

Exclusion of Students:

As indicated in the text, we excluded some of the student observations provided by the universities from the estimation samples used in this report. For the most part, these were records of students that we had not requested and for whom we had no matching records in our OUAC data. The table lists the total number of records provided by the four universities and the number of records excluded for various reasons. A few students had a record in two of our universities. In such cases when the records had different entry years (probable transfers), we kept the record with the earlier entry year. In such cases when the records had the same year, we kept the record for the university at which the student enrolled for more than one year.

The table below indicates that there is a total of 128,166 students in our persistence data set for whom the universities had at least one year of data. The regression reported all have fewer observations because they require that the university had two or more years of data on the student.

Observations Excluded From Data Received from Four Universities		
	Number	Percentage
Total observations received from the four universities.	172143	100.0%
Student entered prior to September 1994.	16896	9.8%
Student could not be matched with 101 OUAC observations. Likely a 105 applicant, e.g., mature or part-time or not Ontario.	10845	6.3%
Student’s postal code in the OUAC application data could not be matched to an Enumeration/Dissemination Area in Ontario.	296	0.2%
Student did not enrol in Level One.	391	0.2%
Student enrolled for fewer than 8 or more than 20 half courses.	9798	5.7%
Student's age upon entry was less than 15 or greater than 20.	1966	1.1%
Key variable is missing.	1369	0.8%
Student had record in two universities. Likely transfer or data error.	90	0.1%
Student is not a September entrant.	2326	1.4%
Number of observations in data set for this report ("Persistence Data Set")	128166	74.5%

Allocation of Students to Programs:

We asked the universities to classify the students in each term by academic program. We asked for both broad and narrow categories. The broadest categories are those used by the Council of Ontario Universities and OUAC. These categories are indicated in the table below. We recognize that some degree of disaggregation by program is desirable given the differences between faculties in entrance standards and academic programs. Our experience with these data indicated that numbers of students were sufficient to allow disaggregation for only the following four categories: Arts, Science, Business and Engineering. Hence, where data were reported in greater disaggregation, we reallocated students in programs other than these four. For confidentiality reasons we cannot give complete details but we give a few examples to illustrate the nature of the reallocations:

Original Program	Reallocated Program
Kinesiology (Physical Education)	Science
Environment Studies	Science
Music	Arts
Nursing	Science
Social Work	Arts

Measures of Academic Progress:

Time Frame for Analysis

The universities reported courses and grades by academic term (fall, winter and summer). One challenge for measuring academic progress is posed by coop programs. All of our universities have coop programs though the degree of student participation varies by institution. Students in coop programs typically are less likely than non-coop students to be enrolled in fall and winter terms (save for the fall of year 1) and are more likely to be enrolled in summer terms. Put differently, course loads are more evenly spread throughout the calendar year for coop students than for non-coop students. Students in coop programs still do take vacation terms and, as a result, progress towards a degree is typically slower than for non-coop students. Unfortunately, the data do not directly indicate which students are in coop programs or which students have no credits in a given term due to being on a coop term. In our multivariate analysis, we control for this to some extent by including control variables for university and program. A second challenge is posed by full-year (two-term) courses which are most common in the first year. Such courses do not show up as passed or failed until the winter term and, as such, can give the mistaken appearance of a light course load when one looks at courses completed in the fall term.

We incorporate the above considerations into our measures of academic progress in two ways. First, we measure such indicators as credits completed and cumulative grade averages on a twelve-month academic year (September-August) basis rather than on an academic term basis. Second, when measuring degree completion rates, we use a wide data window so as to allow for the slower pace of coop students.

Credits Towards Degree

Table 1 contains the measures of academic persistence and success, the first two of which are the cumulative credits passed at the end of 1 and 2 calendar years after entry. By "end of calendar year", we mean the end of the summer term after fall entry. We use a credit measurement system under which 0.5 credits is given for a one-term course and, hence, 5 credits would be the most common one-year, full-time academic load. The sample means of these measures calculated over all four universities and years of data are 4.8 after one year and 9.2 after two years. Note that these two means are based on different sample sizes because we do not have data on two calendar years for all students in our sample. We only have one year of data for the most recent cohorts of students in our samples, two years of the data for the penultimate cohorts and so on. The relationships between the one-year and two-year measures and the characteristics of the individuals, neighbourhoods and high schools in our sample were quite similar. Hence, we focus on the results for cumulative credits passed after two years in the discussion below.

Grade Averages

The universities provided both term and cumulative grade averages. Two of the universities use a 0-100 grading system while the other two use a 0-12 system. We wanted to convert the grades to a single metric to allow analysis across all 4 universities at once. It seemed to make most sense to convert to the 100 point system as it would then be possible to easily consider how an additional grade point average in high school (always on a 100 point system) translated into a different grade point average at university. A big challenge in this regard is the manner in which a grade of F is treated for purposes of calculating grade averages. In the 12 point system an F is always a zero. In the 100 point systems, a grade greater than zero is typically assigned except in cases where the student never showed up or did no work.

We approached this problem in three different ways and confirmed that our regression estimates do not vary from one method to another. Our basic method, and the one we report on in this paper, is to directly translate the 12 point system grades into the 0-100 range according to the following table. The conversion tries to conserve the alphanumeric codes so that, for example, a B in one system remains a B in the other system. Note that we are converting grade point averages so that the grade in both systems (12 point and 100 point) is a continuous variable (GPA* refers to a 100 point system and GPA to a 12 point system).

GPA 12-point system	Conversion Formula
GPA < 0.5	GPA* = 40
0.5 ≤ GPA ≥ 1.0	GPA* = 40 + 24*(GPA-.5)
1.0 < GPA ≥ 2.0	GPA* = 52 + 4*(GPA-1.0)
2.0 < GPA ≥ 3.0	GPA* = 40 + 2*(GP2.0)
3.0 < GPA ≥ 4.0	GPA* = 40 + 4*(GP3.0)
4.0 < GPA ≥ 5.0	GPA* = 40 + 4*(GPA-4.0)
5.0 < GPA ≥ 6.0	GPA* = 40 + 2*(GPA-5.0)
6.0 < GPA ≥ 7.0	GPA* = 40 + 4*(GPA-6.0)
7.0 < GPA ≥ 8.0	GPA* = 40 + 4*(GPA-7.0)
8.0 < GPA ≥ 9.0	GPA* = 40 + 2*(GPA-8.0)
9.0 < GPA ≥ 10.0	GPA* = 40 + 4*(GPA-9.0)
10.0 < GPA ≥ 11.0	GPA* = 40 + 8*(GPA-10.0)
11.0 < GPA ≥ 12.0	GPA* = 40 + 10*(GPA-11.0)

The second approach we employed involved creating a ‘standard normal’ variable for each institution separately by subtracting from each student’s GPA the mean for the institution and dividing by the standard deviation. We then rescaled all 4 distributions so that they had the mean and standard deviation of the two universities (averaged) that started with a 0 to 100 scale. Doing this would be expected to do away with any differences in average marks between institutions and any related individual university effects. If regression estimates for other variables are roughly the same as with the first method, which turn out to be the case, this provides support for the conversion we do use in the paper.

The third method we employ involves working with the equivalent letter grades. Recognizing that in the 100 point system a “B or better” grade is one that is “70 or better” while in the 12 point system a “B or better grade” is represented by marks of “7 or above”, we created a discrete (0, 1) variable that takes on the value of 1 if the student has a B or better average. A similar technique allows one to create a discrete variable for A or better. We then looked at multivariate analysis for the determinants of these discrete variables. The size of the coefficients cannot, of course, be compared directly, but the statistical significance, sign and relative magnitudes of the coefficients can be compared. We concluded from such a comparison that we were not making any incorrect inferences in using our first conversion method.

Departures

We use the term “departure” rather than more commonly used “dropout” because the former more accurately reflects what one can infer from our data. We do not know the destination of student who ceases to register for courses in our four universities. Some departures undoubtedly are voluntary transfers to other universities or colleges or temporary absences from PSE as documented by Finnie and Qiu (2008). Indeed the only means we have of inferring a departure is from courses passed and failed term by term. We note, however, that those students who do depart clearly have below grade averages and credits completed. We count as a “departure during the first year” any student for whom we have no courses passed or failed in the fall, winter or summer of the second academic year. (The number of “departures during the first year” is only slightly lower if we count the students for whom we have no courses passed or failed in the fall, winter or summer of the second and third academic years.) We count as a “departure during the first two years” any student for whom we have no courses passed or failed in the fall, winter or summer of the third academic year.

For each measure, we make sure that our data follow the students for two and three years respectively. That is, we only calculate the one (two) year departure rate among students for whom we have two (three) years of data. We also make sure not to count a student who has earned an early degree as a “departure”. Table 1 indicates that the departure rate during the first year is 8% and during the first two years is 13%. The difference between these two figures is the 5% of entering students who depart during year two.

Degrees Completed

Our final measure of persistence is degree completion. We have information on whether or not a degree is earned, the type of degree earned and the time taken to earn the degree. Our analysis of the data indicated that the distinction between a “General” degree and an “Honours” degree was not that informative. Many “General” degree programs take four full-time years just like “Honours” degrees. Furthermore, some non-Honours programs, such as Engineering, make academic demands of the student that are at least as extensive as Honours degrees in other disciplines. Hence, we did not use this distinction in constructing our measures of degree completion.

Most of the degrees in our data would typically take three or four years to complete if pursued on a full-time basis but there are also five year joint degree programs. More importantly, there are many students in our data who take more than the minimum numbers of full-time years needed to complete a degree program due to coop terms, program switches, periods of part-time study, academic terms abroad, etc. Only 45% of the students in our sample complete any degree within four years. Such early completers are disproportionately from Arts and Science programs and from those universities at which coop programs are less common. The proportions completing a degree within five or six years are 74% and 80% respectively. Such longer data windows also yield a very different pattern of degree completion rates by program and university.

We estimated our multivariate models using each of these measures of degree completion, namely, whether or not a student had completed a degree with four, five or six years of entry. The longer data windows reveal, we believe, a more accurate picture of completion rates by program. In addition, we found that the coefficients for individual, neighbourhood and high school characteristics are quite similar across regressions with different measures of degree completion. For these reasons, we focus in this report on the regression estimates in which the dependent variable is whether or not the student earned any degree within six years after entry. We found this also to be a measure used in other studies of academic success in university (for example see Nora, Barlow and Crisp 2005).

Matching University Data to OUAC Data

OUAC and the universities have provided us with the linking information necessary to allow us to match the OUAC application record to the university administrative record. As expected we are able to find matches in OUAC for almost 100% of the observations provided to us by the universities. The very few exceptions are likely last minute registrants, misclassified (e.g. out-of-province or transfer) students or clerical errors.

Matching University Data to Census OUAC Data

The OUAC data contain the postal code for each applicant at the time of application. We use this to link the student to the Census Enumeration/Dissemination Area (EA/DA) of residence for the student's family. These neighbourhood characteristics provide indicators of the socioeconomic background of the student. Postal codes in the university data are often associated with the student's housing choice at the university and could not be used for this purpose.

There were challenges involved in this process. The first is that there are changes over time in the EA/DA with which a postal code is associated. This is due, most importantly, to the fact that population growth leads over time to the creation of new postal codes and changing postal code boundaries. Fortunately, most postal codes are associated with only one EA/DA in any given calendar year. For the few exceptions to this rule, we chose to identify a postal code with the most recent EA/DA with which that code is associated.

A second challenge arises from the fact that the values of the neighbourhood characteristics are suppressed by Statistics Canada if the sampled population of the EA/DA is too small. Hence, one needs a strategy to find alternate values in cases where there is a missing value for a "key variable". The EA/DA "key variables" for this research project are the following: average household income, average number of persons per household (to be used to adjust average household income for differences in household size), proportion of the population with a BA degree, proportion of the adult population unemployed, proportion of the population that are immigrants, and the proportion of families that are headed by lone mothers. If the EA/DA assigned by Statistics Canada to a postal code has a suppressed value for any key variable, then we search for the nearest postal code that has an assigned EA/DA with a non-missing values for all key variables and use ALL of values from that alternative EA/DA. This means all postal codes have values for key variables from just one EA/DA.