Canadian Journal of Higher Education
Revue canadienne d'enseignement supérieur

The Determinants of Student Success in University: A Generalized Ordered Logit Approach

Philippe Cyrenne and Alan Chan

Volume 52, Number 1, 2022

URI: https://id.erudit.org/iderudit/1089396ar
DOI: https://doi.org/10.47678/cjhe.v52i1.189007

See table of contents

Publisher(s)

Canadian Society for the Study of Higher Education

ISSN

2293-6602 (digital)

Explore this journal

Cite this article

Cyrenne, P. & Chan, A. (2022). The Determinants of Student Success in University: A Generalized Ordered Logit Approach. *Canadian Journal of Higher Education / Revue canadienne d'enseignement supérieur*, *52*(1), 83–107. https://doi.org/10.47678/cjhe.v52i1.189007

Article abstract

The ability of universities and colleges to predict the success of admitted students continues to be a key concern of higher education officials. Apart from a desire to see students have successful academic careers, there is also the fiscal reality of greater tuition revenues providing needed support for university budgets. Using administrative data, this article introduces a relatively new empirical approach to estimating the determinants of student success in post-secondary institutions. Using Ordered Logit and Generalized Ordered Logit estimators, we estimate the role a number of key factors play in influencing student success. As a test of robustness we also use the feologit estimator which is designed to fit fixed effects ordered logit models. An important feature of our approach to determining student success is that it can be conducted using readily avail-able administrative data. While the results are based on one institution, we feel there are useful lessons for other institutions facing similar student performance issues.

THE DETERMINANTS OF STUDENT SUCCESS IN UNIVERSITY: A GENERALIZED ORDERED LOGIT APPROACH

PHILIPPE CYRENNE
UNIVERSITY OF WINNIPEG

ALAN CHAN
GOVERNMENT OF MANITOBA

Abstract

The ability of universities and colleges to predict the success of admitted students continues to be a key concern of higher education officials. Apart from a desire to see students have successful academic careers, there is also the fiscal reality of greater tuition revenues providing needed support for university budgets. Using administrative data, this article introduces a relatively new empirical approach to estimating the determinants of student success in post-secondary institutions. Using Ordered Logit and Generalized Ordered Logit estimators, we estimate the role a number of key factors play in influencing student success. As a test of robustness we also use the *feologit* estimator which is designed to fit fixed effects ordered logit models. An important feature of our approach to determining student success is that it can be conducted using readily available administrative data. While the results are based on one institution, we feel there are useful lessons for other institutions facing similar student performance issues.

Keywords: Generalized Ordered Logit, student success, universities, colleges

Résumé

Pour les responsables de l'enseignement supérieur, la capacité des collèges et universités à prévoir la réussite des étudiants admis demeure une préoccupation constante. Derrière le souhait de voir les étudiants réussir leur parcours d'études, il y a également la réalité financière nécessitant d'accroître les revenus tirés des frais de scolarité afin de soutenir le budget des universités. Pour évaluer les facteurs déterminants de la réussite des étudiants dans les établissements d'enseignement postsecondaires, nous avons introduit dans cet article une approche empirique relativement nouvelle fondée sur des données administratives. Grâce à des estimateurs logit ordonné et logit ordonné généralisé, nous avons évalué l'influence d'un certain nombre de facteurs sur la réussite des étudiants. En guise de test de robustesse, nous avons également utilisé l'estimateur « feologit », conçu pour s'adapter aux modèles logit ordonné à effets fixes. L'aspect important de notre approche est qu'elle peut se fonder sur des données administratives aisément disponibles pour déterminer le succès des étudiants. Et bien que nos résultats ne reposent que sur une seule université, nous pensons qu'ils seront riches d'enseignements pour d'autres établissements confrontés aux mêmes enjeux posés par la réussite des étudiants.

Mots-clés : logit ordonné généralisé, réussite étudiante, universités, collèges

Introduction

The ability of universities and colleges to retain previously admitted students continues to be a concern of higher education officials. Apart from a desire to see students have successful academic careers, there is the fiscal reality of higher tuition revenues providing needed support for university budgets. This desire has spurred a large number of researchers to examine the factors that influence the student performance in higher education, by examining issues related to graduation, stop out and withdrawal.[1]

A number of approaches have been taken to estimate the relationship between student characteristics, institutional support and student performance. One approach is to predict the retention or withdrawal rates of students using *survival analysis*, also called an *event history study*. Survival analysis is the branch of statistics that examines time to a significant event.[2] In higher education, the event might be the time to withdrawal, either permanent or temporary, or time to graduation.[3] The type of survival analyis used depends on the nature and frequency of the data. Continuous time models are based on frequent observations, while discrete time methods are seen as more appropriate for studies of higher education, given that data on post-secondary students is collected on a semester or annual basis. While addressing student performance using discrete time survival analysis can yield helpful insights, the approach requires signficant data preparation as well as the need to address the issue of competing risks regarding student outcomes.[4]

Another approach used to explain the academic outcomes of students are bivariate or multinomial choice models. These models relate the different student outcomes to a number of student characteristics as well as the level of institutional support.[5] A shortcoming of bivariate or multinomial models is they ignore the ranking of alternative student outcomes, for example graduation, continuation, or withdrawal. The fact that multinominal models do not rank outcomes is at odds with the fact that students and university officials have a sense that a student's performance can be ranked with graduation preferable to continuation, which is preferable to withdrawal.[6]

This article builds on the latter approach by examining the factors that determine the relative success of students at one post-secondary institution. Our objectives are two-fold; first to discuss the Generalized Ordered Logit estimator, and second to apply it to a unique data set to identify the key determinants of student success at a mid size liberal arts institution. The central research question is to determine the factors that influence the graduation, continuation, or withdrawal of previously admitted students. In doing so, we develop a relatively novel approach that, as far as we know, has not been widely used in the higher education literature. Specifically, we use both an Ordered Logit and Generalized Ordered Logit model to determine the factors that can lead to greater success in university or college.[7] In defining relative success, we assume the student outcomes can be ranked ordinally, with graduation the highest ranked outcome, followed by continued enrolment, with permanent withdrawal as the lowest ranked outcome.

In order to address the issue of the factors influencing student success using a number of alternative estimators, we use a unique data set that includes information on the subsequent performance at the University of Winnipeg of students from 83 Manitoba high schools, drawn from 31 Manitoba school divisions. By tracking their performance over time, we determine the likelihood of success of students enrolled at the University of Winnipeg, based on their initial academic performance as well as a number of characteristics, including their high school grades, the nature of their high school education, and financial support.

Our article differs from previous work in several ways. First, we track students from several entering classes over a five year period.[8] This allows us to record whether students graduated, continued or withdrew from the university within that period.[9] Second, our data set only involves students that graduated from Manitoba high schools, which are governed by a province-wide curricula and employ provincially certified teachers.[10] Third, our data set includes a significant number of covariates that can assist us in identifying the types of students that may be more successful.[11] In addition, our data allows us to examine the effect of a student's high school grades, the nature of the student's high school (whether private or public), as well as its funding level, on post-secondary success. Fourth, by using an Ordered Logit Model and Generalized Ordered Logit model, we develop a simple framework for evaluating the determinants of students success using commonly collected administrative data.[12] We feel the Generalized Ordered Logit model is a flexible estimator that can be easily applied by unversity officials even in settings where student data is limited.[13] Fifth, in forecasting student success, we only use data on the respective covariates one year after admittance.[14] Finally, we examine the determinants of success for several student types, which include visa students, students who were granted permanent resident status in Canada, and students self-identified as Aboriginal.[15]

Our findings are the following. First, we find a student's first-year performance is a key indicator of student success. We also find that students pursuing an education degree, students entering with better high school grades, or students who took a larger number of full course equivalents in their first year were more successful. We find little evidence of significant high school fixed effects on student success. In addition, we find that the parallel lines/proportional odds assumption assumed in the standard Ordered Logit model does not hold for several covariates. To account for this violation, we re-estimate the model using the Generalized Ordered Logit model as discussed by Williams (2006, 2016). We find the odds ratios associated with several regressors differ significantly with respect to the different categories of success. While our results apply to one university, we feel the results can be applied to a wide range of institutions that share similar academic structures and similar information gathering.[16]

A Brief Literature Review

A number of researchers have examined the factors that influence the retention of college students. Using data on full time community college freshmen, Dey and Astin (1993) find few practical differences between analyses conducted using logit, probit, or linear regression. They find the strongest predictor of retention was a student's high school average, with concern about finances, a desire to make more money, and preparation for graduate school also playing a significant role.[17]

Montmarquette et al. (2001), using a bivariate probit model with selectivity bias, find that strong academic performance in the first semester, as well as a smaller class size in first year mandatory courses, help explain the persistence of the university students. Singell (2004) uses a random utility approach and data on in-state and out-of-state freshmen to examine whether financial aid affects college retention. Singell finds that while need- and merit-based financial aid increase retention, increasing reliance on unsubsidized and merit-based aid by the government and universities lowers the relative graduation rates of needy students.[18]

Herzog (2005), using a multinominal logit model, finds that middle-income students with financial challenges are more at risk of dropping out, while first-year math experience, second-year financial aid, and simultaneous enrolment at another college or university are important determinants of student retention. In contrast, Arulampalam et al. (2005), using a binomial logit model and a large data set of UK undergraduate students, find that weaker students are more likely to drop out during their first year of study. They also find that that better prepared students are less likely to leave highly ranked universities, while weaker students at the same universities face significant pressure to do so.

Stratton et al. (2008), using a multinomial logit random utility model and data from the National Center for Education Statistics, find that delayed matriculation, the nature of first-year aid received, and marital and personal status have a significant effect on whether a student's academic career is interrupted. They find that a substantial fraction of withdrawals are temporary, and the nature of the financial aid has a significant effect on whether a student continues or drops out.

Danilowicz-Gösele et al. (2017) use a Probit model and administrative data from a German university and find that student success, defined as graduation, differs significantly by faculty. They find a student's entering grade is a strong predictor of academic success, while measures of parental income or social background do not add significantly to the explanatory power of the model.[19] Similarly, Heck et al. (2012) examine the determinants of student success using an ordered logit model with the outcomes ordered from drop out to continued enrollment to graduation from high school.[20] Using data from 6,883 students from 934 high schools, they find that lower levels of socioeconomic status or high absenteeism raise the log of the odds of dropping out or remaining enrolled in high school (Heck et al., 2012, p. 298).

Regarding our policy contribution, while examining higher education issues using administrative data is not new, a goal of our research is to help university officials address retention issues. While estimators like survival analysis can help institutions identify students at risk, they require significant data preparation and relatively highly skilled administrative staff to implement. In contrast, we feel the Generalized Ordered Logit approach requires less sophistication in terms of data collection, data preparation, and model estimation. In addition, we argue a Generalized Ordered Logit approach can help predict which type of students are at risk, despite only using data from a student's first year university performance. We feel this reduced administrative burden, as well as model parsimony, can be useful to higher education administrators concerned with student retention and performance.

Possible Hypotheses

Along with much of the literature, we assume that a student's high school grades are a strong predictor of university performance (Cyrenne & Chan, 2012). Regarding gender, recent academic research suggests that in recent years, females are outperforming their male counterparts, both at the high school and post-secondary level (Cohn et al., 2004). We suspect that certain types of students may face obtacles in their pursuit of higher education. For example, Richardson and Blanchet-Cohen (2000) provide a comprehensive analysis of the challenges facing Aboriginal students and discuss a number of issues that need to be addressed to support Aboriginal students in post-secondary education programs.

Regarding the high school effect, it is possible the subsequent performance of Manitoba high school students is affected by the resources spent on their high school education (*rexpenditure*), as well as the non-pecuniary features of the high school (academic standards, discipline).[21] For example, a common perception is that students from private high schools perform better at the University of Winnipeg quite apart from differences in high school resources.[22] We separate these two effects by recording both school expenditures and whether the student graduated from a public or private school, and whether the private school was religious-based or secular.[23] Financial variables, which can include family income support, bursaries, and scholarships, are viewed as very important in helping students succeed in their post-secondary studies.[24] However, given the University of Winnipeg does not record the student's financial background upon admittance, we use as a proxy the median family income of the postal code that is listed as the student's permanent address (*rincome*). We also include information regarding the choice of major by students. For example, we use dummy variables to indicate whether a student did not declare a major (*deqnoinfo*) or were admitted to the Education program prior to enrolling at the University of Winnipeg (*deqedu*).[25]

The Estimation Procedure

The first step in the estimation procedure is to define what is meant by student success in higher education. We assume the highest ranked outcome is graduation (coded as 3), followed by continued enrolment (coded as 2), and finally withdrawal as the lowest ranked outcome (coded as 1), all within the five-year period.[26]

In order to estimate the determinants of the above rankings of student outcomes we use both an Ordered Logit Model and a Generalized Ordered Logit model. An Ordered Logit model is designed to estimate the the underlying score (the ranking of outcome) as a linear function of the independent variables and a set of cutpoints. Specifically, "the probability of observing outcome i corresponds to the probability that the estimated linear function, plus random error, is within the range of the cutpoints estimated for the outcome."[27]

As outlined by Williams (2006), the standard Ordinal Logit model (*ologit*) can be written:

$$P(Y_i > j) = g(X\beta) = \frac{\exp(\alpha_j + X_i\beta)}{[1 + \{\exp(\alpha_j + X_i\beta)\}]}, \quad j = 1,2, \ldots\ldots, M-1$$

The standard model is sometimes called the parallel lines/proportional odds model.[28]

The Generalized Ordered Logit model (*gologit*) we feel is computationally feasible and yields improved estimates of the marginal probability effects over the standard model, which can be written as

$$P(Y_i > j) = g(X\beta_j) = \frac{\exp(\alpha_j + X_i\beta_j)}{[1 + \{\exp(\alpha_j + X_i\beta_j)\}]}, \quad j = 1,2, \ldots\ldots, M-1$$

As can be seen, the difference is that in the standard Ordinal Logit model, the β's are constrained to be equal across categories, that is, for each value of *j*.[29] As suggested by Williams (2006), the parallel lines assumption, or fixed β's across categories, is often violated. In the case where the β's vary across categories, the probabilities that the dependent variable *Y* will take on the respective values, 1,…,*M* are given as

$$P(Y_i = 1) = 1 - g(X_i\beta_1)$$

$$P(Y_i = j) = g(X_i\beta_{j-1}) - g(X_i\beta_j) \qquad j = 2, \ldots\ldots, M-1$$

$$P(Y_i = M) = 1 - g(X_i\beta_{M-1})$$

Depending on the number of categories defined as M, when M = 2 we get the logistic regression model; with M > 2 as outlined by Williams, the model becomes a series of binary logistic regessions.[30] The Generalized Ordered Logit can be estimated using the estimator *gologit2* written for Stata.[31]

Before presenting the results from the Generalized Ordered Logit estimator (*gologit2*), it is helpful to discuss the meaning of the empirical results in terms of the odds ratios with respect to the covariates. As discussed by Liu (2016), the Generalized Ordinal Logit model estimates the cumulative odds of being beyond a certain category versus being at or below that category. Specifically,

$$Odds(Y > j) = \frac{p(Y>j)}{p(Y<j)}$$

In our case, with the three categories, category 3 is a student who graduated within five years, category 2 is a student who is still registered in year five, and category 1 is if the student has withdrawn by year five (did not graduate and was no longer registered).

For example, regarding the odds a student is above category 1, that is, is no longer registered is

$$Odds(Y > 1) = \frac{p(Y>1)}{(1-p(Y>1))} = \frac{(p(2)+p(3))}{p(1)} \qquad (1)$$

where *p*(*j*) is the probability of category *j*. Similarly, the odds that a student is above category 2, that is still registered is

$$Odds(Y > 2) = \frac{p(Y>2)}{(1-p(Y>2)} = \frac{p(3)}{p(1)+p(2)} \tag{2}$$

Or in terms of the category names

$$Odds(Y > withdrawn) = \frac{p(Y>withdrawn)}{(1-p(Y>withdrawn))} = \frac{(p(continuing)+p(graduating))}{p(withdrawn)} \tag{1'}$$

$$Odds(Y > continuing) = \frac{p(Y>continuing)}{(1-p(Y>continuing))} = \frac{p(graduating)}{(p(continuing)+p(withdrawn)} \tag{2'}$$

As a test for robustness, we also used a relatively new estimator, *feologit*, developed by Baetschmann et al. (2020), which is designed to estimate fixed effects ordered logit models. Given that our model does not have panel data, but high school fixed effects, our discussion of the details of the *feologit* estimator will be brief.[32]

The *feologit* estimator is designed to estimate causal effects by allowing for some flexibility with respect to unobserved time-invariant heterogeneity, which is often present in fixed effect models.[33] The *feologit* estimator has two options, the BUC estimator (blow up and cluster) and the BUC-τ, which assumes constant thresholds. These estimators are based largely on the conditional maximum likelihood estimator (CML), which is implemented in Stata as the *clogit* and the panel-data command *xtlogit*, *fe* commands.

The *feologit* estimator was developed to address the issue that while a consistent fixed effects estimator exists for the binary logit model, finding a consistent fixed effects estimator for the ordinal logit model has proved challenging. Chamberlain (1980) was able to find a consistent estimator by collapsing the dependent variable into a binary variable and then applying the conditional maximum likelihood (CML) estimator. The approach taken by Baetschmann et al. (2016) is based on extending the Chamberlain approach by considering a larger number of dichotimizations. The approach involves including more than one "clone" of an individual combined with cluster standard errors as used by the BUC estimator (Baetschmann et al., 2011) to estimate the parameter vector β in the ordered logit model, fixed effects model with individual specific thresholds, and then, given that the clones of each other are not independent of each other, the standard errors are based on clustering at the individual level. Creating these clones expands the number of observations used in estimating the parameter β.

The Data Set

The data involves a cross section of student cohorts who entered the University of Winnipeg over a five-year period. The first cohort in our sample entered the University of Winnipeg in 1997.[34] The data set was created by one of the authors from merging separate data sets while employed in the Institutional Analysis Department of the University of Winnipeg.[35] The University of Winnipeg is a liberal arts institution, with relatively liberal admission standards, a significant range in student preparation, but a relatively strong faculty overall. The University of Winnipeg is a primarily undergraduate institution, which in many ways is similar in structure and mission to four-year public colleges found in the United States. It is funded in much the same way as state colleges in the United States, with the majority of operating funds coming from the Province of Manitoba.

During the period we tracked the performance of students, the University of Winnipeg consisted of three faculties: the Faculty of Education, the Faculty of Science, and the Faculty of Arts (including Humanities). Unlike most universities, the University of Winnipeg did not have a Faculty of Business at the time, though one was created in 2008, several years after our sample period. Prior to that, business studies consisted largely of a set of recommended courses from a wide variety of departments supplemented with a small set of business courses.

It is important to outline how we arrived at the final data set.[36] Our final data set includes 5,008 observations. Our initial sample included 14,246 observations; however, a number of students were dropped from this sample for a variety of reasons. First, only students who graduated from a Manitoba high school were included.[37] Second, a number of students were dropped because they did not have a standard high school average.[38] Third, some adult learners (older than 21) who did not graduate from high school (Manitoba or otherwise) but were admitted to the University of Winnipeg under a Mature Student category were dropped. Fourth, students for which family income data could not be estimated, or were missing school expenditure data from their respective high schools, were dropped.[39] Finally, students who did not have a five-year average GPA at the University of Winnipeg were excluded from the sample. The above exclusions left us with 5,187 observations, while data limitations regarding students from a number of smaller high high schools reduced the data set to 5,136 observations.[40] We further excluded 128 students who dropped out completely prior to the end of the first year, which resulted in 5,008 students in the final data set. Our data set tracked the subsequent performance of the entering classes of 1997 through to 2001. Once admitted, we tracked the course registrations and university grades for each entering class over a five-year period.

Table 1 provides summary statistics on our sample of 5,008 students from 83 Manitoba high schools. Table 1 shows that the majority of students, 97%, were classified as Canadian (*canadian*), with males making up approximately 36% of the students (*male*). The mean age of students was 18.8 years (*age*). Overall, the high school average (*hsaverage*) of incoming students over the five-year period (1997–2002) was 78.38%. We also indicate whether a student was enrolled at the University of Winnipeg on a student visa (*visa*; 77 students), was a permanent resident in Canada (permanent; 86 students), or self-identified as Aboriginal (*aboriginal*; 93 students).[41]

We also recorded the academic plans of students, with 116 students pursuing a four-year degree (*deg4yr*), 813 an Education degree (*degedu*), while 2,383 students had not declared their degree plans (*deqnoinfo*).[42] In addition, we recorded the grade point average of students after their first year (*gpayr1*) and the number of full course–equivalent courses students completed after their first year (*fcesyr1*). The mean GPA of students after year one was 2.77 amongst those who completed 3.54 full course–equivalent courses in their first year on average. Also included are dummy variables to record whether a student received financial assistance. Overall, 4% of students received either a University of Winnipeg (*dUWbursary1*) or Province of Manitoba bursary (*dMBbursary1*), while 29% of students received a University of Winnipeg scholarship (*dUWscholarship1*) and 11% of students received a loan from the Province of Manitoba (*dMBloan1*).[43]

For each student we recorded the high school from which the student graduated.[44] The sample includes students from 83 Manitoba high schools, with the top 10 high schools sending slightly more than 50% (2,584) of the total number of students. The students in our sample came from 76 public schools and seven private schools. Private high schools are assigned to the Funded Independent School Division and receive some public funding from the Province of Manitoba.[45] Regarding private high schools (*private*), they include a mix of secular and religious-based high schools (*religprivate*). In terms of the mix of students, 18% of the students graduated from a private high school, while 10% graduated from a private high school that was religious-based.

We also collected school expenditures per pupil determined at the school division level (*rexpenditure*).[46] The 5,008 students came from 31 school divisions with 4,705 students or 94% of the students coming from the top 10 divisions. During the sample period, the largest number of first year enrolments came from the Winnipeg School Division (969) followed by Funded Independent Schools (882). The remaining students came from several school divisions in the city of Winnipeg and a large number of rural school divisions. There is a substantial variation in expenditure by school divisision, with the lowest real expenditure per student being $3,436 while the highest is $13,840, with a median real expenditure of $5,921.

While data on student finances would be ideal, the University of Winnipeg does not record the financial background of students or the educational level of parents or guardians in their admission process. Given educational attainment is often viewed as related to family income, we use as a proxy for family income, the median family income associated with the respective postal code given as their permanent residence (*rincome*).[47] It addition, the variable *rincome* can capture any peer or neighborhood effects.

Table 2 summarizes the performance of the 1997–2002 entering classes at the University of Winnipeg. Of the 5,008 students, 1,814 graduated within a five-year period, while 3,194 did not.[48] The withdrawal rate is very high, ranging from a low of 40% for the class of 2000 to a high of 50% for the students admitted in 1997. The percentage of students graduating within five years ranged from a low of 32% for the 1997 entering class to a high of 39% for the 1999 class. Much like the trend for other post-secondary institutions, a significant number of students are taking longer to complete their degrees, with 19% of students continuing their registration after five years.

Table 3 reports the summary statistics for three subsamples of our data set, which include self-identified Aboriginal students, permanent residents and visa students.[49] There is some variation in the summary statistics for the three groups; however, there are some general trends. In terms of incoming high school average (*hsaverage*), first-year GPA (*gpayr1*), and student success, Aboriginal students tend to be quite similar to visa students. Visa students are more likely to have attended a private high school than Aboriginal students or permanent residents. Students who are permanent residents tend to have a higher incoming high school average, a higher first-year GPA, take more courses in their first year, and come from higher income neighborhoods than visa or Aboriginal students. Aboriginal students tend to be somewhat older, more likely to pursue an education degree, and more likely to be female than visa or permanent resident students.

Regarding financial support, 15% of self-identified Aboriginal students received a loan from the Province of Manitoba, 13% received a bursary from the Province of Manitoba, and 13% received a scholarship from the University of Winnipeg. In contrast, 30% of permanent residents received a loan from the Province of Manitoba, 22% received a University of Winnipeg scholarship and 8% a University of Winnipeg bursary. Regarding visa students, 4% did receive a scholarship from the University of Winnipeg.[50]

Ordered Logit Estimation Results

Table 4 summarizes the results based on equation (1). For each specification, we report both the logit coefficients and the corresponding odds ratios. Regarding the odds ratio, a value greater than one suggests the covariate has a positive effect on the odds of success for a particular category, a value less than one suggests a negative effect on the respective odds, and an odds ratio equal to one implies no relationship between the particular covariate and the respective odds.

Column 1 of Table 4 reports the logit coefficients and odds ratios for the Ordered Logit estimator excluding high school fixed effects, while column (2) reports the corresponding estimates including the high school fixed effects. For column (2) we dropped the regressors, *private* and *rexpenditure*, since they were correlated with the school fixed effects. The results are very similar for the latter two specifications. We find little evidence of high school fixed effects contributing to student success, as only six high school dummy variables were statistically significant.[51] As a test of robustness, we also present estimates using the *feologit* estimator, designed to fit fixed effects ordered logit models. Column (4) uses the BUC estimator, while Column (3) uses the BUC-τ, which assumes contant thresholds.

To summarize our results, students attending the University of Winnipeg on a student *Visa*, pursuing an education degree (*deqedu*), entering with a higher high school average (*hsaverage*), taking more courses in their first year (*fceyr1*), or earning a higher first-year university GPA (*gpayr1*) had greater success at the University of Winnipeg. In contrast, students who received a loan from the Province of Manitoba (*dMBloan1*) were less successful. This is an interesting result which could reflect the fact that students requiring a loan from the Province of Manitoba may be in a more precarious financial situation. The other measures of student aid were not statistically significant.[52] As a test of robustness we included two interactive terms (*drelig_private=religious x dprivate*), which test whether attending a religious private high school effected success, whether education students entering with higher high school averages were more successful than non-education students (*degedu_hsavg=deqedu x hsavg*), as well as an age squared term (*agesq*). Given these effects were not statistically significant for specifications (1) to (2), they were dropped from the models.[53]

Regarding the overall fit of the Ordered Logit models, we report two measures: the McFadden Pseudo R2, which is 0.094 for model (1) and 0.104 for model (2), and the McKelvey and Zavoina R2 measures, which are 0.202 and 0.224 respectively.[54] Veall and Zimmermann (1996) argue the McKelvey and Zavoina R2 has a number of desirable properties as a measure of goodness of fit for a number of limited dependent variable models.[55] Overall, the McKelvey and Zavoina measures for each model suggest the respective models are relatively good predictors of student success.[56]

As can be seen, the results for the feologit estimator given in columns (3) and (4) generally mirror the results using the Ordered Logit estimator with high school fixed effects given by column (2), with similar parameter estimates and similar levels of statistical signifance.

While it is often customary to report the marginal effects associated with the Ordered Logit Model, we prefer to focus on the appropriateness of the parallel lines/proportional odds assumption that underlies the Ordered Logit Model.[57]

Ordered Logit Results – Testing for the Parallel Regression Assumption

Regarding the Ordered Logit Results, a key assumption is the parallel lines/proportional odds assumption, which implies stability of the β coeffficients across categories. The Brant test, the results of which are listed in Table 5, formally tests this restriction. Table 5 reports the coefficient estimates that result from a series of binary regressions across the categories. The Brant test of the parallel lines/proportional odds assumption is based on a chi-sq test, a significant test statistic that provides evidence that the assumption has been violated. As can be seen the respective coefficients, the β_j vary significantly for the regressors *dvisa*, *degedu*, *gpayr1*, *fceyr1*, and *rincome*.[58]

Generalized Ordered Logit Results (gologit2)

Table 6 lists the Generalized Ordered Logit results for the model excluding high school fixed effects. We report both the logit coefficients as well as the corresponding odds ratios. As the results indicate, the Brant tests suggest that restrictions should be imposed on the β's for 17 of the 24 regressors, which includes restrictions on four time fixed effects. This results in identical coefficients for those regressors which are omitted from Table 6 for ease of presentation.[59] In addition, once these restrictions are imposed, the Wald test of the parallel lines or proportional odds for the model yields a $\chi^2(17) = 16.25$, which yields the prob>$\chi^2 = 0.5065$. As listed in the Stata output, an insignificant test statistic suggests the final model does not violate the parallel lines/ proportional odds assumption. The McFadden Pseudo R2 is a respectable 0.1108.[60]

As pointed out by Liu (2016, p. 191), it is important to discuss the regressors that meet the parallel lines/proportional odds assumption as well as those that violate the assumption. The regressors that meet the parallel lines/proportional odds assumption include the 16 regressors that are constrained in the model. The coefficients of the regressors (β) that indicate the immigration status of the student (*dvisa* or *dpermanent*), the choice of major (*degedu*), the student's first-year performance (*gpayr1*), course load (*fcesyr1*), financial resources (*rincome*) and high school resources (*rexpenditure*), and the (*y2002*) entering class are not constrained and are allowed to vary across the categories.

Overall, four of the regressors are statistically significant across the respective categories, with the respective odds of being above the respective categories of success increasing for students with a higher first-year GPA (*gpayr1*), students who took more first-year courses (*fcesyr1*), and visa students (*dvisa*), but decreasing for declared education students (*degedu*).

The results as summarized as follows. The odds ratios reflect the probability of being above a category. Regarding the odds ratios for category 1 – withdrawn, there are 6 regressors that are statistically significant—*visa*, *degedu*, *gpayr*, *fcesyr1*, *dMBloan*, and *rincome*—with increases in all these covariates except *dMBloan*, leading to a greater probability of a student continuing, that is, being above category 1, withdrawn.

Regarding the odds ratios for category 2 – continuing, only 4 covariates are statistically significant. The covariates which increase odds of graduating are (i) being a visa student (2.69), (ii) being enrolled in education (1.63), (iii) having a higher first-year GPA (1.75), and (iv) taking more full-course equivalents in the first year (1.82).

Most of these results accord with intuition. Factors such as taking a larger number of full-course equivalents, more success in the first year, and being a visa student raise the probability of both continuing and graduating. Each of these covariates are monotonic across categories, that is, increases in the levels raise the probability of graduating more than the probability of continuing. This might suggest that there is a momentum effect for students, those students that place ever more value on continuing, and hence even more important to graduate given their success in first year.[61]

Of interest is that for education majors (*degedu*), the odds of graduating are 2.4 times, and the odds of continuing are 1.6 times, the comparable odds for non-education majors. Of note is that unlike the other covariates, this pattern is not monotonic,[62] that is, the odds of continuing for education majors are relatively higher when compared to non-education majors than the respective odds for graduating.[63]

Discussion

This article has highlighted two issues related to the determinants of student success at a mid-sized Canadian university. The first examines the use of an Ordered Logit estimator to address the issue of student success. Given the proportional odds assumption is often violated when using the Ordered Logit estimator, this leads naturally to the Generalized Ordered Logit estimator. This estimator allows for a test of the proportional odds assumption well as the imposition of a set of contraints on the ordered logit coefficients in order to satisfy the proportional odds assumption. We think the Generalized Ordered Logit Estimator will gain in popularity given its flexibility and relative ease of interpretation and implementation.

The second issue is the application of these estimators to a particular undergraduate university. The data set was compiled by one of the authors and provides a comprehensive examination of the relationship between the student characteristics and their academic success. The Ordered Logit and Generalized Ordered Logit are flexible in that a wide variety of what constitutes success can be estimated. Our measure suggests that students who continue their studies are more successful over a five-year observation period than students who withdraw.

It is important to note there are varying measures of student success that can be considered depending on the respective goals of university officials and policy makers. For example, it might be that some students who do not complete their degrees in a timely manner should reconsider their career plans. From the perspective of university officials, students who continue their studies contribute to university budgets; however, from the perspective of policy makers some of these students may be better suited to other types of training.[64]

The results from this study can be summarized as follows. First, we find, much like other studies, that students who achieve a higher first-year GPA are more likely to succeed. Second, we find that students who take a larger number of full-course equivalents in first year are more successful.[65] Third, we find visa students and students enrolled in education programs are more successful while students who received a loan from the Manitoba government were less successful. We find little evidence of high school fixed effects. In testing the robustness of the Ordered Logit results, we find the parallel lines/proportional odds assumption is violated for a number of covariates. To address this issue, we used a Generalized Ordered Logit estimator, which imposes equality of coefficients on those covariates that satisfy the parallel lines/proportional odds assumption, while allowing the other covariates to vary across the categories of success.

While the results are based on data from one institution, we think our approach can assist other institutions in designing their data collection methods. We feel our study is a relatively straightforward approach to predicting the success of a variety of students using data that is routinely collected by university officials. We urge university institutional research departments to set up a template to record the characteristics of incoming students, which can then be updated with student performance measures in order to use an Ordered Logit or Generalized Order Logit estimator. For example, in developing a database to use an Ordered Logit or Generalized Ordered Logit estimator, a number of programming issues need to be addressed. First, data on incoming students is often kept separate from databases that track their performance. To overcome this issue, this study merged the respective information using Structure Query Language procedure which is a computer language used to store, manipulate, and query data stored in relational databases. This allowed us to link the data sets together through common identifiers such student ID, school name, and postal code. We also recommend that universities collect as much background information on incoming students as possible as allowed by privacy legislation. If comprehensive data is collected by universities at the admission stage and stored in a central database, it is relatively easy to implement the estimators used here to determine the likelihood of student success.

Disclaimer

The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the University of Winnipeg or the Government of Manitoba

References

Allison, P. D. (2014). *Event history and survival analysis* (2nd ed.). Sage.

Arulampalam, W., Naylor, R. A., & Smith, J. P. (2005). Effects of in-class variation and student rank on the probability of withdrawal: Cross-section and time-series analysis for UK university students. *Economics of Education Review*, *24*(3), 251–262. https://doi.org/10.1016/j.econedurev.2004.05.007

Baetschmann, G. (2012). Identification and estimation of thresholds in the fixed effects ordered logit models. *Economics Letters*, *115*(3), 416–418. https://doi.org/10.1016/j.econlet.2011.12.100

Baetschmann, G., Ballantyne, A., Staub, K. E., & Winkelmann, R. (2020). feologit: A new command for fitting fixed-effects ordered logit models. *The Stata Journal*, *20*(2), 253–2275. https://doi.org/10.1177%2F1536867X20930984

Baetschmann, G., Staub, K. E., & Studer, R. (2016). Does the stork deliver happiness? Parenthood and life satisfaction. *Journal of Economic Behavior & Organization*, *130*, 242–2260. https://doi.org/10.1016/j.jebo.2016.07.021

Baetschmann, G., Staub, K. E., & Winkelmann, R. (2011). *Consistent estimation of the fixed effects ordered logit model* (IZA Discussion Paper No. 5443). The Institute for the Study of Labor. https://ftp.iza.org/dp5443.pdf

Boes, S., & Winkelmann, R. (2006). Ordered response models. *Allgemeines Statistisches Archiv—Journal of the German Statistical Society*, *90*, 167–181.

Box-Steffensmeier, J. M., & Jones, B. S. (2004). *Event history modeling: A guide for social scientists*. Cambridge University Press.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, *47*(1), 225–238. https://doi.org/10.2307/2297110

Chen, R. (2008). Financial aid and student dropout in higher education: A heterogeneous research approach. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (pp. 209–239). Springer.

Cohn, E., Cohn, S., Balch, D. C., & Bradley, J., Jr. (2004). Determinants of undergraduate GPAs: SAT scores, high-school GPA and high-school rank. *Economics of Education Review*, *23*(6), 577–586. https://doi.org/10.1016/j.econedurev.2004.01.001

Coleman, J. S., & Hoffer, T. (1987). *Public and private high schools: The impact of communities*. Basic Books.

Cyrenne, P., & Chan, A. (2012). High school grades and university performance: A case study. *Economics of Education Review*, *31*(5), 524–542. https://doi.org/10.1016/j.econedurev.2012.03.005

Danilowicz-Gösele, K., Lerche, K., Meya, J., & Schwager, R. (2017). Determinants of students' success at university. *Education Economics*, *25*(5), 513–532. https://doi.org/10.1080/09645292.2017.1305329

DesJardins, S. L., Ahlburg, D. A., & McCall, B. P. (1999). An event history model of student departure. *Economics of Education Review*, *18*(3), 375–390. https://doi.org/10.1016/S0272-7757(98)00049-1

DesJardins, S. L., Ahlburg, D. A., & McCall, B. P. (2002). A temporal investigation of factors related to timely degree completion. *The Journal of Higher Education*, *73*(5), 555–581. https://doi.org/10.1080/00221546.2002.11777168

DesJardins, S. L., Ahlburg, D. A., & McCall, B. P. (2006). The effects of interrupted enrollment on graduation from college: Racial, income, and ability differences. *Economics of Education Review*, *25*(6), 575–590. https://doi.org/10.1016/j.econedurev.2005.06.002

Dey, E. L., & Astin, A. W. (1993). Statistical alternatives for studying college student retention: A comparative analysis of logit, probit, and linear regression. *Research in Higher Education*, *34*(5), 569–581. https://doi.org/10.1007/BF00991920

Ermisch, J., & Francesconi, M. (2001). Family matters: impacts of family background on educational attainments. *Economica*, *68*(270), 137–156. https://doi.org/10.1111/1468-0335.00239

Fu, V. (1998). Estimating generalized ordered logit models. *Stata Technical Bulletin*, *8*(44), 27–29.

Gallop, C. J., & Bastien, N. (2016). Supporting success: Aboriginal students in higher education. *Canadian Journal of Higher Education*, *46*(2), 206–224. https://doi.org/10.47678/cjhe.v46i2.184772

Greene, W. H., & Hensher, D. A. (2010). *Modeling Ordered choices: A primer*. Cambridge University Press.

Gross, J. P. K., Hossler, D., Ziskin, M., & Berry, M. S. (2015). Institutional merit-based aid and student departure: A longitudinal analysis. *The Review* *of Higher Education*, *38*(2), 221–250. http://doi.org/10.1353/rhe.2015.0002

Hagle, T. M., & Mitchell, G. E., II. (1992). Goodness-of-fit measures for probit and logit. *American Journal of Political Science*, *36*(3), 762–784. https://doi.org/10.2307/2111590

Häkkinen, I., Kirjavainen, T., & Uusitalo, R. (2003), School resources and student achievement revisited: New evidence from panel data. *Economics of Education Review*, *22*(3), 329–335. https://doi.org/10.1016/S0272-7757(02)00060-2

Hanushek, E. (1997). Assessing the effects of school resources on student performance: an update. *Educational Evaluation and Policy Analysis*, *19*(2), 141–164. https://journals.sagepub.com/doi/pdf/10.3102/01623737019002141

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2012). *Multilevel modeling of categorical outcomes using IBM SPSS*. Routledge.

Herzog, S. (2005). Measuring determinants of student return vs. dropout/stopout vs transfer: A first-to-second year analysis of new freshman. *Research in Higher Education*, *46*, 883–928. https://doi.org/10.1007/s11162-005-6933-7

Horowitz, J. B., & Spector, L. (2005). Is there a difference between private and public education on college performance? *Economics of Education Review*, *24*(2), 189–195. https://doi.org/10.1016/j.econedurev.2004.03.007

Hosmer, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis: Regression modeling of time-to-event data*. Wiley.

Kennedy, P. (1998). *A guide to econometrics* (4th ed.). MIT Press.

Liu, X. (2016). *Applied ordinal logistic regression using stata*. Sage.

Long, J. S., & Freese, J. (2014). *Regression models for categorical dependent variables using stata* (3rd ed.). Stata Press.

Montmarquette, C., Mahseredjian, S., & Houle, R. (2001). The determinants of university dropouts: A bivariate probability model with sample selection. *Economics of Education Review*, *20*(5), 475–484. https://doi.org/10.1016/S0272-7757(00)00029-7

Murtaugh, P. A., Burns, L. D., & Schuster, J. (1999). Predicting the retention of university students. *Research in Higher Education*, *40*(3), 355–371. https://doi.org/10.1023/A:1018755201899

Neal, D. (1997). The effects of Catholic secondary schooling on educational attainment. *Journal of Labour Economics*, *15*(1), 98–123. https://www.jstor.org/stable/2535316

Ott, M. D., Markewich, T. S., & Ochsner, N. L. (1984). Logit analysis of graduate student retention. *Research in Higher Education*, *21*(4), 439–460. http://www.jstor.org/stable/40195665

Richardson, C., & Blanchet-Cohen, N. (2000). Postsecondary education programs for Aboriginal peoples: Achievements and issues. *Canadian Journal of Native Education*, *24*(2), 169–184.

Singell, L. D., Jr. (2004). Come and stay a while: Does financial aid effect retention conditioned on enrollment at a large public university? *Economics of Education Review*, *23*(5), 459–471. https://doi.org/10.1016/j.econedurev.2003.10.006

Smith, J., & Naylor, R. (2005). Schooling effects on subsequent university performance: Evidence for the UK university population. *Economics of Education Review*, *24*(5), 549–562. https://doi.org/10.1016/j.econedurev.2004.07.016

St. John, E. P., & Noell, J. (1989). The effects of student financial aid on access to higher education: An analysis of progress with special consideration of minority enrollment. *Research in Higher Education*, *30*(6), 563–581. https://www.jstor.org/stable/40195922

Stratton, L. S., O'Toole, D. M., & Wetzel, J. W. (2008). A multinomial logit model of college stopout and dropout behavior. *Economics of Education Review*, *27*(3), 319–331. https://doi.org/10.1016/j.econedurev.2007.04.003

Tinto, V. (2006). Research and practice of student retention: What next? *Journal of College Student Retention*, *8*(1), 1–19. https://doi.org/10.2190%2F4YNU-4TMB-22DJ-AN4W

Veall, M. R., & Zimmermann, K. F. (1996). Pseudo-R2 measures for some common limited dependent variable models. *Journal of Economic Surveys*, *10*(3), 241–259. https://doi.org/10.1111/j.1467-6419.1996.tb00013.x

Verbeek, M. (2012). *A guide to modern econometrics* (4th ed.). John Wiley & Sons.

Webber, D. A., & Ehrenberg, R. G. (2010). Do expenditures other than instructional expenditures affect graduation and persistence rates in American higher education? *Economics of Education Review*, *29*(6), 947–958. https://doi.org/10.1016/j.econedurev.2010.04.006

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *The Stata Journal*, *6*(1), 58–82. https://doi.org/10.1177%2F1536867X0600600104

Williams, R. (2016). Understanding and interpreting generalized ordered logit models. *The Journal of Mathematical Sociology*, *40*(1), 7–20. https://doi.org/10.1080/0022250X.2015.1112384

Windmeijer, F. A. G. (1995). Goodness-of-fit measures in binary choice models. *Econometric Reviews*, *14*(1), 101–116. https://doi.org/10.1080/07474939508800306

Wolfe, R., &, Gould, W. (1998). An approximate likelihood-ratio test for ordinal response models. *Stata Technical Bulletin*, *7*(42), 24–27.

Wooldridge, J. W. (2013). *Introductory econometrics: A modern approach* (5th ed.). South-Western.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *The Review of Economics and Statistics*, *85*(1), 9–23. https://doi.org/10.1162/003465303762687677

Contact Information

Philippe Cyrenne
p.cyrenne@uwinnipeg.ca

Notes

1 For a nice summary of the literature on student retention up to 2006, see Tinto (2006).

2 For good introductions to event study analysis, see Box-Steffensmeier et al. (2004), Hosmer et al. (2008), and Allison (2014). In the medical literature, it examines the factors that influence the time to a heart attack.

3 The higher education literature refers to permanent withdrawal as dropout, temporary withdrawal as stopout. Articles using the event history approach are Murtaugh et al. (1999), DesJardins et al. (1999, 2002, 2006), and Herzog (2005).

4 By competing risks we mean that the occurrence of one event removes the individual from the risk of the occurrence of another event. In the higher education context, a student drop out removes the individual from the the hazard of graduating. Similarly, the occurrence of dropping out for one year may affect the hazard of permanently dropping out. While according to Allison (2014, p. 57), at the theoretical level including these competing risks do not introduce any new issues, at a practical level it introduces more flexibility in estimating survival analysis models. It is our view the presence of competing risks raises considerably the analytical burden for practitioners, particularly those not well versed in survival analysis. See Allison (2014) for a good introduction to discrete time event study analysis, also called discrete time survival analysis.

5 See, for example, Montmarquette et al. (2001), Herzog (2005), Arulampalam et al. (2005), and Stratton et al. (2008).

6 In the higher education literature, there is a distinction between the permanent withdrawal of a student and what is a temporary withdrawal, often referred to as a stopout. See Herzog (2005) for a model that makes that distinction. There is also a distinction made between withdrawal from a university and a transfer to another post-secondary institution.

7 As a test of robustness, we also use the ordered logit estimator for panel data, feologit, introduced by Baetschmann et al. (2015).

8 The five-year observation period was a result of data availability. It is clear that a longer observation period may reveal that some students took longer than five years to graduate.

9 Unfortunately, like other studies using administrative data for a single institution, we are not able to track students who withdraw from one university to attend another post-secondary institution. In a sense, our withdrawal category includes both students who leave the University of Winnipeg to transfer to another post-secondary institu-

tion or drop out of university entirely.

10 This allows us to control for the relationship between a student's high school curriculum and subsequent performance in university, which in turn allows us to control for this unobserveable school effect, an issue which often plagues studies of higher education.

11 For example, we include whether students received financial assistance, either a bursary or a scholarship, while attending the University of Winnipeg.

12 In contrast event study analysis for higher education depends on discrete time methods, which requires a careful collection and organization of data to be useful for empirical work. See Allison (2004) for a good discussion of data requirements for discrete time event study analysis.

13 In contrast, discrete time survival methods require a continuous updating and recording of student performance as well as other time varying covariates, a significant administrative burden.

14 While adding information on student performance and institutional support in subsequent years may be useful determinants of graduation and continuing enrolment, it makes the analysis of the determinants of withdrawal problematic. For instance, some students withdraw after the first year, which makes the inclusion of second- and third-year covariates inappropriate for this group.

15 In Canada, the designation of First Nations people has changed over time, moving from the term Aboriginal to First Nations, with the present designation being Indigenous peoples. We use the term Aboriginal, which was the term used when the data was collected. For an article that discusses a number of ways of describing success with respect to Aboriginal students in higher education, see Gallop and Bastien (2016).

16 In particular, the results we feel are representative of liberal arts colleges and universities, which have relatively liberal admittance criteria and relatively strong faculty. It may seem surprising that a university with relatively liberal admittance standards could have a relatively strong faculty; however, Canadian universities receive significant government support, which allows Canadian universities to offer relatively attractive faculty salaries and working conditions as well as a relatively large number of places for students in Canadian universities.

17 See Ott et al. (1984) for a logit model of graduate student retention.

18 For a nice summary of a number of theoretical approaches and issues that address the relationship between financial aid and student dropout in higher education, see Chen (2008).

19 See Cyrenne and Chan (2012) for an analysis of the relationship between a student's high school grades and subsequent university performance using administrative data for a Canadian university.

20 In should be noted we only became aware of this research after our empirical work was completed.

21 One important issue that we are not able to address with our data set is the effect of peer effects on university success. See Zimmerman (2003) for an article that addresses this issue.

22 See, for example, Horowitz and Spector (2005) for an analysis of the relative difference in performance of students graduating from private and public high schools. Smith and Naylor (2005) find that the students who attended an independent school in the United Kingdom were less likely to obtain a "good' degree than students who attended a state-sector school. Hanuskek (1997) and Häkkinen (2003) also examine this issue.

23 There is significant literature on the issue of the relative effects of religious education; for example, Coleman and Hoffer (1987) and Neal (1997).

24 There is evidence that the educational attainment of children is strongly influenced by the educational attainment of parents, as argued by Ermisch and Francesconi (2001). Unfortunately, the University of Winnipeg does not record the education level of parents for students who are admitted. However, we do have an estimate of family income, which is, in general, highly correlated with educational attainment. See Gross et al. (2015) for an analysis of the relation of merit-based aid to student departure.

25 We examine two hypotheses regarding the relative success of Education students. The first is that students in professional programs like Education are more committed to their degree program. The second is that graduating with an Education degree may be relatively easier than for other programs. For both these reasons, it is suggested that students enrolled in Education are more likely to succeed.

26 Regarding outcome 3, we look to see if the student was enrolled in year five. If not, and the student had not graduated in the five-year period, the student is coded 3, withdrawn. Given we are only able to observe students over a five-year period, we recognize that it is possible that some students who had not graduated or registered in the fifth year may in fact re-enroll in a later period. It is clear there could be a larger set of ordinal outcomes, involving a finer partition of states. For example, we could distinguish between students who graduated within four years from those who graduated in five years. While this may be of interest to some observers, we feel an expansion of the event space will make the interpretation of the empirical results somewhat unwieldy.

27 STATA Release 12 Reference N-R (2011:1413).

28 As discussed by Boes and Winkelmann (2006), the standard ordered response model has a number of limita-

tions when analyzing marginal probability effects, which includes a constant threshold assumption while the distributional assumption does not allow for additional individual heterogeneity between different realizations. Boes and Winkelmann (2006) discuss a number of versions of the generalized ordered response model and point out each of them can be computationally burdensome due to the larger number of parameters to be estimated and the larger number restrictions on the parameter space.

29 A nice intuitive illustration of the generalized ordered logit model as compared to the standard ordered logit, parallel lines model is Williams (2016). The parallel lines model is sometimes called the proportional odds model, see Wolfe and Gould (1968) who use this terminology.

30 Verbeek (2012, p. 222) discusses the alternative normalizations that can occur for σ and μ_1. See Kennedy (1998) for a discussion of the ordered probit and ordered logit estimators. In general, the results from the ordered probit and ordered logit estimators tend to be very similar. See also Long and Freese (2006) for a nice overview of the estimation of ordered categorical dependent variables using STATA.

31 It has been suggested that gologit2 was inspired by Fu's (1998) gologit program. We use Stata 12's ologit to estimate the Ordered Logit Model and gologit2 to estimate the Generalized Ordered Logit Model. For an interesting analysis of Generalized Ordered models, see Greene and Hensher (2010), Chapter 6.

32 A nice overview of the discussion of the feologit approach, which uses the BUC (blow up and cluster) approach, is Baetschmann (2012). As can be seen there, the intuition given for the estimation is based on panel data; that is, multiple ordered observations of a given individual.

33 For a good discussion of the feologit estimator see Baetschmann et al. (2020) and Baetschmann et al. (2015). Articles that provide insight into the background of the estimator are Baetschmann et al. (2011), Baetschmann (2012), and Baetschmann et al. (2016).

34 The first cohort of students was classified as a first-time student who successfully passed at least one course at the University of Winnipeg.

35 As an employee of the Institutional Analysis Department of the University of Winnipeg, Alan Chan was able to access student data from a number of different data sets. The use of the administrative data has been approved by the University of Winnipeg Ethics Committee. While it would be interesting to update the data set, we feel the current data set provides a good sample to address the issues examined in this article.

36 It is important that the sample be representative of the population of students in order to avoid sample selection bias.

37 It is clear there are a number of possible research questions related to our question. For example, with sufficient data, one could include all students who attended the University of Winnipeg (not just from Manitoba high school students) and estimate the associated high school effects for those students. Alternatively, we could compare the performance of Manitoba high school students at the University of Winnipeg with their performance at other post-secondary institutions. Apart from the issue of insufficient data for these exercises, the first question would face the issue of varying high school curricula across jurisdictions, while the second would need to control for varying grading standards across post-secondary institutions.

38 For example, some students who might have a letter grade for Grade 12 (or equivalent) courses rather than a numerical score.

39 For some students their address (that is, their postal code) did not match Statistics Canada records with respect to family income (using the first three digits of their postal code).

40 Regarding the issue of the smaller high schools, 13 high schools only sent one student, and 17 high schools sent two students for a total of 47 students from 30 high schools being excluded from the sample. At the time the data was being collected it was felt that these small numbers would not adequately reflect the nature of the student's high school and would in effect reflect a student fixed effect (at least for the 13 high schools who sent one student). We were not able to collect all the necessary regressors for the remaining students (from schools who sent two students), so they were dropped from the sample.

41 These 93 students are included in the 4,845 total designated as Canadian. We break foreign students into two categories: permanent resident and visa students. It can be argued that permanent residents (as a result of receiving landed immigrant status) and visa students may differ in terms of academic behaviour and performance. For instance, permanent residents may have more social engagement than visa students and may have more flexibility regarding the number of courses taken than visa students. On the other hand, permanent residents who are allowed to work may spend more time working at part-time jobs. On the contrary, visa students may try to speed up their academic programs, sacrificing their overall performance. Finally, permanent residents may have more opportunities to improve their English language skills than visa students.

42 It is important to note that, unlike many other universities in Canada or the United States, the University of Winnipeg has a three-year degree program for Arts, Science,

and Business.

43 To be eligible for the University of Winnipeg's General Bursary (*UWBursary1*) a student must prove financial need and be making satisfactory academic progress (for example maintaining a "C" average). Applicants are asked to estimate their expenses and resources; the latter includes savings as well as any monies received from parents and others, income of spouse or partner, and part-time employment income over the fall/winter term, as well as loans and awards. As is the case with bursaries in general, students do not have pay back amounts received. Regarding University of Winnipeg Scholarship (*UWScholarship1*) over the period of 1997–1999, the high school average cut-off rate to be awarded an entrance scholarship was 84%, while after 2001 it was reduced to 80% for recruitment purposes.

44 We realize that high school students often change high schools prior to graduation; however, we only have data on the student's high school of graduation.

45 It is important to note that all schools in Manitoba, both publicly funded and private, are subject to the same curriculum guidelines, outlined by the Manitoba Education, Citizenship and Youth Department of the Province of Manitoba.

46 This was based on data from Manitoba Education, Citizen and Youth. All the school expenditure and average family income were calculated as the real term (nominal value divided by Manitoba Education Price Index; 1996=100) and were merged into the data set. This is expenditure at the school division level. Webber and Ehrenberg (2005) find that the type of college expenditure affects graduation rates. In particular, enhanced expenditure on student services influences both graduation and first-year persistence rates, particular for schools with lower entrance test scores.

47 There are two possible interpretations of our measure of family income. The first is that this variable is a proxy for the variable of interest, family income. The second is that it captures a neighbourhood effect. Our study adopts the first interpretation, which is based on the idea that family income is highly correlated with the median family income of the postal code in which the student resides. To estimate a student's family income we used the median income from the postal code listed as the student's permanent residence (for each year, based on the first three digits of the postal code).

48 It should be pointed out that students who did not graduate may not have dropped out of university permanently, as some students in this group enrolled in another university, or took longer than five years to graduate. Unfortunately, our data set does not capture the respective size of these effects.

49 Permanent residents are students who have received landed immigrant status in Canada.

50 Visa students are not eligible for financial support from the Province of Manitoba.

51 It should also be noted that for a number of high schools there are only a few students who are in our sample.

52 With respect to the student aid effect, there is some evidence that the impact of aid may vary depending on the type of student. St. John and Noell (1989) find that, at least when it comes to college attendance, all forms of student aid had a stronger impact on minority student access to college than for non-minority students. Given the relatively small number of minority students—in our case, Aboriginal students—testing for this effect would have low statistical power. However, it is important to note that some self-identified Aboriginal students, particularly First Nations students, do receive some support from the First Nation Band council for post-secondary studies. Unfortunately, our data set does not include that information.

53 As discussed by Long and Freese (2014), it can happen that one of the regressors predicts the highest category perfectly. In contrast, if the regressor predicts the middle category perfectly, there is no estimation problem. In the former case, Long and Freese suggest the problematic observations be dropped from the model.

54 A number of Pseudo R2 measures have been proposed. For a nice summary of the relative merits of these measures see Veall and Zimmermann (1996).

55 Similarly, Long and Freese (2014) report that Hagle and Mitchell (1992) and Windmeijer (1995) using simulations find that for ordinal outcomes the McKelvey and Zavoina's R^2 measure closely approximates the R^2 obtained from a linear regression on the respective ordinal outcomes.

56 It is also possible to include data from the second and third year of the student's performance. As might be expected, the predictive ability of the generalized ordered logit estimator is greatly increased. However, given that a significant number of students drop out after the first year, the sample size is reduced significantly.

57 The marginal effects indicate the effect of a unit increase in a variable, either continuous or a dummy variable, on the probability of the respective outcome. Some marginal effects are worth reporting, for example, regarding Outcome 3 (graduating) taking one more fce in year one (*fceyr1*) increases the probability of a student graduating by 8.8%, one point higher first-year GPA (*gpayr1*) increases the probability of graduating by 9.3%, while Education students are 12.6% more likely to graduate, and visa students are 13.9% are more likely. It is important to note that students who receive a loan from the Manitoba government are 6.7% less likely to graduate.

58 We have only reported the coefficients from the Brant detail test, for ease of exposition. The accompanying standard errors with respect to the coefficients are available on request.

59 One thing that must be acknowledged is the following: Warning! 8 in sample cases have an outcome with a predicted probability that is less than 0. See the gologit2 help section on Warning Messages for more information.

60 The GOLOGIT2 estimator does not report the McKelvey and Zavoina statistic. Other measures that are reported, for example, Cox-Snell/ML (0.207) and Cragg-Uhler/Nagelkerke (0.236) suggest a relatively good fit for the gologit2 model.

61 There may also be an element of learning by doing for students—those that successfully navigated first year may develop study and learning skills that allow them to build on their initial success.

62 I thank an anonymous referee for pointing out the relative performance of Education students.

63 Two possible explanations come to mind. First, given the large retention in Education majors after first year, it may be the case that some of these Education students switch to other majors; that is, they do not graduate with an Education degree. Second, given the large first-year retention, it may be the case that some of these Education students do not complete their Education degree in the five-year time frame. Note that both of these explanations are based on the relative performance of Education students over time who are still continuing and graduating at a higher rate than non-Education majors. These are potential research questions to be explored.

64 In conclusion, what constitutes success for post-secondary students is clearly an issue that can benefit from further discussion.

65 This may be a measure of full-time status, which could reflect the fact that these students have greater financial support to pursue their studies.

Appendix 1
Definition of Variables

Variable	Definition
success	1 = withdrawn, 2 = continuing, 3 = graduating
canadian	Canadian citizen
permanent	Permanent resident
visa	Visa student
aboriginal	Self-declared Aboriginal student
deg4yrs	4-year major declared
degedu	Education major declared
degnoinfo	Undeclared major
private	Graduated from private high school
religprivate	Graduated from religious-based private high school
dUWbursary1	Received UW bursary in year 1
dUWscholarship1	Received UW scholarship year 1
dMbloan1	Received loan from Province of Manitoba year 1
dMBbursary1	Received Province of Manitoba bursary year 1
hsaverage	High School Average (%)
gpayr1	Grade Point Average after year 1
fcesyr1	Full-course equivalent courses completed after year 1
age	Age when first admitted
rincome (000's)	Real income of postal code of student's permanent residence
rexpenditure (000's)	Real expenditure of school division in which the high school is located.

Table 1

Summary Statistics – Incoming Students (5,008)

variable	total	mean	sd	min	median	Max
success		1.92	0.8943	1	2	3
canadian	4845	0.97	0.1775	0	1	1
permanent	86	0.02	0.1299	0	0	1
visa	77	0.02	0.1231	0	0	1
aboriginal	93	0.02	0.1350	0	0	1
male	1792	0.36	0.4794	0	0	1
deg4yr	116	0.02	0.1504	0	0	1
degedu	813	0.16	0.3688	0	0	1
degnoinfo	2383	0.48	0.4995	0	0	1
private	882	0.18	0.3810	0	0	1
religprivate	500	0.10	0.2998	0	0	1
dUWbursary1	108	0.02	0.1453	0	0	1
dUWscholarship1	1439	0.29	0.4526	0	0	1
dMbloan1	550	0.11	0.3127	0	0	1
dMBbursary1	123	0.02	0.1548	0	0	1
hsaverage (%)		78.38	10.0231	51	79	100
gpayr1		2.77	0.9164	0	2.75	4.5
fcesyr1		3.54	1.2491	0.5	4	7.5
age		18.83	2.4811	16	18	69
real income (000's)		48.833	13.2971	18.400	48.744	91.999
real school expenditure (000's)		6.077	0.9035	3.436	5.921	13.840

Source: University of Winnipeg Administrative Data, Statistics Canada, Province of Manitoba

Table 2

Status of the 5,008 Students (Five Years After Entry)

	Ordered Outcomes - Success			
cohort	1	2	3	
	Withdrawn	Continuing	Graduated	Total
1997	367	131	242	740
1998	380	138	285	803
1999	317	150	306	773
2000	326	183	299	808
2001	371	183	337	891
2002	463	185	345	993
Total	2,224	970	1,814	5,008

Source: University of Winnipeg Administrative Data

Table 3

Summary Statistics - Sub Samples

	success	male	deg4yr	degedu	degnoinfo	private	dMBloan1	dMBbursary1	dUWbursary1	dUWscholar1	hsaverage	gpayr1	fcesyr1	age	rincome	rexpenditure
Aboriginal (*N* = 93)																
sum		20	1	19	38	14	14	12	2	12						
mean	1.83	0.22	0.01	0.20	0.41	0.15	0.15	0.13	0.02	0.13	73.81	2.36	3.27	21.23	41.911	6.186
min	1	0	0	0	0	0	0	0	0	0	55	0	0.5	17	18.983	4.148
median	2	0	0	0	0	0	0	0	0	0	75	2.4	3	19	43.338	6.042
max	3	1	1	1	1	1	1	1	1	1	94.67	4.25	5.5	54	89.337	8.675
Permanent (*N* = 86)																
sum		40	2	5	42	14	26	3	7	19						
mean	1.70	0.47	0.02	0.06	0.49	0.16	0.30	0.04	0.08	0.22	78.36	2.61	3.40	20.04	44.352	6.219
min	1	0	0	0	0	0	0	0	0	0	57.67	0	0.5	16	18.983	3.957
median	1	0	0	0	0	0	0	0	0	0	79	2.5	3.5	19	46.588	5.988
max	3	1	1	1	1	1	1	1	1	1	96.67	4.5	6	43	91.999	10.543
Visa (*N* = 77)																
sum		37	3	1	38	46	0	0	0	3						
mean	1.90	0.48	0.04	0.01	0.49	0.60	0	0	0	0.04	71.99	2.23	2.71	20.04	36.370	5.705
min	1	0	0	0	0	0	0	0	0	0	52	0	0.5	17	18.400	3.522
median	2	0	0	0	0	1	0	0	0	0	71.33	2.17	2.5	19	37.263	5.912
max	3	1	1	1	1	1	0	0	0	1	89	4.25	7.5	32	53.789	11.143

Table 4

Ordered Logit and Feologit Results

Success	olbase		olschool		feologit (BUC-T)		feologit (BUC)	
	(1)		(2)		(3)		(4)	
Variable	Logit Coeffi-cients	Odds Ratios	Logit Coeffi-cients	Odds Ratios	Logit Coeffi-cients	Odds Ratios	Logit Coeffi-cients	Odds Ratios
Age	-0.0057	0.99	0.0029	1.00	0.0032	1.00	0.0039	1.00
Aboriginal	0.1484	1.16	0.1999	1.22	0.248	1.28	0.2291	1.26
Permanent	-0.2791	0.76	-0.2677	0.77	-0.2783	0.76	-0.2383	0.79
Visa	0.7006**	2.02**	0.5575*	1.75*	0.5880*	1.80*	0.5933*	1.81*
Male	0.0446	1.05	0.0545	1.06	0.0495	1.05	0.0511	1.05
deg4yrs	0.1829	1.20	0.2059	1.23	0.1929	1.21	0.236	1.27
Degedu	0.6327***	1.88***	0.6730***	1.96***	0.7090***	2.03***	0.7118***	2.04***
Degnoinfo	-0.0606	0.94	-0.0486	0.95	-0.0464	0.95	-0.0542	0.95
Hsaverage	0.0079	1.01	0.0104*	1.01*	0.0095	1.01	0.0102*	1.01*
gpayr1	0.4689***	1.60***	0.4686***	1.60***	0.4876***	1.63***	0.4882***	1.63***
fcesyr1	0.4419***	1.56***	0.4502***	1.57***	0.4631***	1.59***	0.4637***	1.59***
Private	-0.0964	0.91						
dMBloan1	-0.3387***	0.71***	-0.3348**	0.72**	-0.3537**	0.70**	-0.3420***	0.71***
dMBbursary1	-0.2285	0.80	-0.2309	0.79	-0.2313	0.79	-0.25	0.78
dUWbursary1	-0.021	0.98	-0.038	0.96	-0.0567	0.94	-0.0375	0.96
dUWscholarship1	0.0298	1.03	0.0139	1.01	-0.0183	0.98	-0.0217	0.98
Rincome	-0.0016	1.00	-0.0027	1.00	-0.003	1.00	-0.0024	1.00
Rexpenditure	0.0038	1.00	-0.0783	0.92	-0.0993	0.91	-0.0963	0.91
yr1998	0.0804	1.08	0.0653	1.07	0.0907	1.09	0.0798	1.08
yr1999	0.2868*	1.33*	0.2828*	1.33*	0.3022*	1.35*	0.3017*	1.35*
yr2000	0.2452*	1.28*	0.2479*	1.28*	0.2751*	1.32*	0.2664*	1.31*
yr2001	0.2412*	1.27*	0.2438*	1.28*	0.2582*	1.29*	0.2656*	1.30*
yr2002	0.0964	1.10	0.1231	1.13	0.1328	1.14	0.1435	1.15
HS Fixed Effects			yes	yes	yes	yes	yes	yes
cut1								
_cons	3.2378***	25.4768***	3.4112***	30.3023***	0	1		

Success	olbase		olschool		feologit (BUC-τ)		feologit (BUC)	
	(1)		(2)		(3)		(4)	
Variable	Logit Coeffi-cients	Odds Ratios	Logit Coeffi-cients	Odds Ratios	Logit Coeffi-cients	Odds Ratios	Logit Coeffi-cients	Odds Ratios
cut2								
_cons	4.1753***	65.0612***	4.3647***	78.6294***	0.9499***	2.5855***		
Statistics								
N	5008		5008		60014		10002	
Ll	-4745.7		-4694.8939		-33038.82		-5498.56	
Aic	9541.4001		9601.7878		66123.63		11041.12	
Bic	9704.3699		10292.7797		66330.69		11199.7519	
McFadden PseudoR2	0.0904		0.104		0.1478		0.1278	
McKel-Zavoina R2	0.202		0.224					

legend: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Table 5

Brant Tests

	Estimated Coefficients from j-1 binary regressions		Test of Parallel Regression Assumption		
Variable	y>1	y>2	chi2	p>chi2	df
All			178.48	0.000	23
age	-0.004	-0.012	0.25	0.614	1
aboriginal	0.174	0.131	0.03	0.854	1
permanent	-0.426	-0.045	4.18	0.041	1
visa	0.479	1.13	11.04	0.001	1
male	0.074	0.001	1.48	0.224	1
deg4yrs	0.267	0.175	0.22	0.642	1
degedu	0.895	0.519	15.33	0.000	1
degnoinfo	-0.067	-0.059	0.01	0.915	1
hsaverage	0.008	0.007	0.07	0.794	1
gpayr1	0.419	0.595	16.18	0.000	1
fcesyr1	0.364	0.593	63.24	0.000	1
private	-0.115	-0.13	0.05	0.827	1
dMBloan1	-0.275	-0.436	2.24	0.134	1
dMBbursary1	-0.402	-0.04	3.25	0.071	1
dUWbursary1	0.117	-0.102	1.12	0.290	1
dUWscholar-ship1	-0.028	-0.015	0.02	0.880	1
rincome	-0.005	0.002	13.18	0.000	1
rexpenditure	0.009	-0.023	1.04	0.308	1
yr1998	0.077	0.112	0.13	0.719	1
yr1999	0.335	0.271	0.33	0.563	1
yr2000	0.349	0.174	2.48	0.115	1
yr2001	0.306	0.22	0.63	0.427	1
yr2002	0.135	0.095	0.14	0.709	1
_cons	-2.858	-4.938			

Table 6

Generalized Ordered Logit Estimates

success	Category 1 – Withdrawn		Category 2 - Continuing	
Variable	Coefficients	Odds Ratios	Coefficients	Odds Ratios
age	-0.0078	0.99		
aboriginal	0.1478	1.16		
permanent	-0.4194	0.66	-0.0302	0.97
visa	0.5028*	1.65*	0.9912***	2.69***
male	0.0425	1.04		
deg4yrs	0.2009	1.22		
degedu	0.8910***	2.44***	0.4873***	1.63***
degnoinfo	-0.0533	0.95		
hsaverage	0.0071	1.01		
gpayr1	0.4125***	1.51***	0.5567***	1.74***
fcesyr1	0.3606***	1.43***	0.5981***	1.82***
private	-0.1095	0.90		
dMBloan1	-0.3268**	0.72**		
dMBbursary1	-0.2177	0.80		
dUWbursary1	-0.0049	1.00		
dUWscholarship1	-0.0004	1.00		
rincome	-0.0047*	1.00*	0.002	1.00
rexpenditure	0.0002	1.00		
yr1998	0.0792	1.08		
yr1999	0.2784*	1.32*		
yr2000	0.3271**	1.39**	0.1362	1.15
yr2001	0.2410*	1.27*		
yr2002	0.0914	1.10		
_cons	-2.5878***	0.08***	-5.0643***	0.01***
Statistics				
N		5008		
ll		-4657.3542		

success	Category 1 – Withdrawn		Category 2 - Continuing	
Variable	Coefficients	Odds Ratios	Coefficients	Odds Ratios
aic		9378.7084		
bic		9587.3098		
Wald chi2(30)		986.88		
Prob > chi2		0.0000		
Pseudo R2		0.1111		

legend: * $p < 0.05$; ** $p < 0.01$; ***$p < 0.001$