High School Grades and University Performance:
A Case Study

Philippe Cyrenne and Alan Chan

THE UNIVERSITY OF WINNIPEG

Department of Economics

515 Portage Avenue

Winnipeg, R3B 2E9

Canada

High School Grades and University Performance: A Case Study

Philippe Cyrenne and Alan Chan*

May 13, 2010

Abstract

Abstract. A critical issue facing a number of colleges and universities is how to allocate first year places to incoming students. The decision to admit students if often based on a number of factors, but a key statistic is a student's high school grades. This paper reports on a case study of the subsequent performance at the University of Winnipeg of high school students from 84 Manitoba High Schools. By tracking the University performance of a set of students admitted for the years 1997-2002, we are able to estimate the likelihood of success of subsequent students based on their characteristics as well as their high school grades. In doing so, we use a number of alternative estimators including a Least Squares Dummy Variable Model and a Hierarchical Linear Model. The methodology should be of interest to admissions officers at other universities as an input into estimating the subsequent performance of first year students.

(JEL L1,L2,L4,L83)

1 Introduction

The issue of the admission policies of universities has attracted considerable attention in recent

years. For example, in the United States there have been a number of legal challenges involving

affirmative action admission policies in a number of states. The elimination of Grade 13 in Ontario brought the issue to the forefront in Canada, where the usual anxiety among parents and students regarding the securing of a university placement had been heightened. Given the increased demands by the health care sector for government funds, it is likely that universities throughout Canada will be forced to look more closely at the issue of rationing as the capacity of universities to admit students becomes problematic. It is possible that other jurisdictions will be forced to make similar choices given increased relative demands on governments as well as the increasing cost of providing higher education in their countries.

This paper reports on some research that has been undertaken at the University of Winnipeg in order to address these emerging issues. A key research question is to determine to what extent conventional admissions data, like high school performance, is helpful in predicting the performance of admitted students. Unlike the U.S., standardized tests such as the SAT, are not in general use in Canada, which means that universities in Canada are much more dependent on student characteristics, including the student's high school grades in making admission decisions. In order to address this issue, data was collected on the subsequent performance at the University of Winnipeg of high school students from 35 Manitoba School Divisions which includes some 84 Manitoba high schools. By tracking the university performance of students who graduated from Manitoba high schools over the years 1997-2002, we are able to determine the likelihood of success of subsequent students based on their characteristics including their high school performance and the nature of their high school education.

A number of researchers have examined the factors that affect the success of students in college

or university. For example, Betts and Morell (1999) address the issue of the performance of high school students in university using a sample of 5,000 undergraduates at the University of California, San Diego. They find that the personal background of students strongly affects their grade point averages (GPA) as undergraduates, and in particular, they find significant high school effects. In measuring the effect of the quality of a student's high school, they find that the experience level of the high school teachers has a positive but small effect on their university GPA. However, Betts and Morell find that neither the teacher-pupil ratio nor the teacher's education level have a statistically significant effect on a student's subsequent university GPA.

More recently, Cohn et al. (2004) assess the degree to which SAT scores, high school GPA and class rank, predict the college GPA of students enrolled at the University of South Carolina. Among their objectives is to determine who is most likely to benefit from statewide college scholarship programs in South Carolina. They find that including an SAT requirement for scholarships is important since it increases the probability of success in college. They also find differences in student achievement depending on race and gender.

Finally, Grove and Wasserman (2004) have examined the life-cycle pattern of student performance for a set undergraduate student cohorts at a large private university in the northeast over a five year period. They found that student grades followed a "check-mark" pattern, with student grades falling after the second semester and rising thereafter, with a slump in the last academic term. They conclude that attrition and participation in the Greek system account for over half of the longitudinal change in academic achievement.

Our paper differs from previous work in this area in a number of ways. First, the focus of our

research is to examine the usefulness of high school grades as a predictor of subsequent university performance. Second, given the above, a related objective is to determine whether there are significant variations in grading standards between high schools. Given that all the students in our sample have graduated from Manitoba high schools, which are governed by a Province wide mandated curriculum, we feel our data set provides the basis for a "natural experiment" in that it allows us to abstract from possible variations in high school curricula across jurisdictions. Third, our data set allows us to examine the effect of high school grades on the life-cycle GPA of students. That is, we are able to determine the effect of high school performance, as measured by high school grades, on the subsequent GPA of students at various stages of their university career. In contrast to Groves and Wasserman (2004), we observe student performance at the end of each academic year, in contrast to their results which apply to student performance on a semester by semester basis. Much of our focus, however, is on the long run performance of students since the variance of first year grades may be significantly higher than the long term performance of students, our results may be helpful to university officials who may be more interested in the long run performance of students. Finally, in contrast to much of the earlier work, we use a variety of estimation approaches, including Least Squares Dummy Variable (LSDV) and Hierarchical Linear Model (HLM) models. The latter approach allows us to determine whether our results are robust to various assumptions regarding the data generating process, specifically, the fact that students are nested within high schools.

Our results can be summarized as follows. First we find significant school effects; however, we find that neither the expenditure per student by the school division or our measure of family

4

income have statistically significant effects on the student's subsequent university GPA. We discuss reasons for this result in the paper. Similar to Groves and Wasserman (2004) we find the effect of high school grades on a student's university performance to diminish over time. Second, we find the LSDV and the HLM estimators yields similar results. For example, we find the performance of Foreign classified students to be significantly lower than the performance of Canadian students, despite the fact that all attended Manitoba high schools. We also find that there is a significant private school effect, with students attending private high schools earning approximately .10 GPA points higher than their public school counterparts. However, we find that the religious based nature of the private school to be statistically insignificant.

While the issue of what should be the key determinants of a student's admission to university may be controversial, we feel the methodology that we develop here provides a benchmark to be used in any decision where the high school grades of potential students is a key admission criteria. As well, the empirical methodology should be of interest to admissions officers at post secondary institutions if part of their selection criteria includes an allocation of first year positions in an efficient manner.

2   The Data Set

The data used in this paper involves a cross section of student cohorts who entered the University of Winnipeg over a five year period. The first set of students in our sample entered the University of Winnipeg in 1997.[1] Once admitted, we tracked the course registrations and university grades

---

[1]The first cohort of student was classified as a first time student who successfully passed at least one course at the University of Winnipeg.

for the 1997 entering class over a five year period. This data collection process was then repeated for the subsequent entering classes of 1998, 1999, 2000, 2001.

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 over which we tracked the performance of students, the University of Winnipeg academic structure consisted of 3 principal 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, which was only created in 2008, several years after our sample period. Prior to that, business studies consisted largely of a set of courses recommended from a wide variety of departments, and was largely viewed as an Interdisciplinary program in the Social Sciences, with a relatively small set of explicitly business type courses in existence.

It is important to outline how we arrived at the final data set.[2] In total we have 5,136 observations in our final data set. There were 14,246 observations in our initial sample; however, a number of students were dropped from this initial sample for a variety of reasons. First, we only considered students who graduated from a Manitoba high school: this included both Canadian and Foreign students.[3] Second, a number of students were dropped because they did not have standard high

---

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

[3]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 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

school marks - i.e. students who might have a letter grade for Grade 12 (or equivalent) courses rather than a numerical score. 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 were ruled out.[4] 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 5187 observations, but a decision to restrict the sample by dropping high schools that sent fewer than 3 students to the University of Winnipeg, resulted in 5136 observations in the final data set. Table 1 summarizes how the final data set was obtained.

[Table 1 here]

Table 2 provides summary statistics on our sample of 5136 students from 84 Manitoba High Schools. In terms of the mix of incoming students, Table 2 shows that the majority of students were classified as Canadian, roughly 97%. Females make up approximately 64% of the students in our sample with the majority of students graduating from Public Schools (82%). The mean age of students was 18.84 years. Overall, the High School Average (HS_{AVG}) of incoming students over the 5 year period (1997-2002) was 78.3, while the mean University Grade Point Average (U_{GPA}) of those students after 5 years was 2.89. In terms of HS_{AVG}, the mean high school grade for females was higher than males, (79.5 versus 76.1) while the mean high school grade for Canadian students was higher than for Foreign students (78.4 versus 75.3) The grades of incoming Private school

standards across post secondary institutions.

[4]For some students their address (that is, their postal code) did not match Statistics Canada records with respect to family income (using the 1st 3 digits of their postal code)

students were not significantly different than the high school grades for Public school students, (78.9 versus 78.2).

[Table 2 here]

In terms of the University Grade Point Average U_{GPA}, the mean grade of females was significantly higher than that for males, (2.96 versus 2.77). The mean U_{GPA} for Canadian students was significantly higher than that of Foreign students (2.91 versus 2.56), while the U_{GPA} of private school students was significantly higher than for students from the public school system (3.02 versus 2.87). It is important to note that our sample of 5136 observations includes both students who graduated over the five year period (1821) and those that did not (3315). Of those that did not graduate, the respective High school marks were lower 76.33 versus 81.88 for students who graduated. As expected, the U_{GPA} after 5 years was also lower, 2.66 versus 3.3004 for those that did not graduate.[5]

Table 3 outlines data on the means of the variables by year of entry for incoming students at the University of Winnipeg over the sample period. As can be seen, there has been increase in the number of first year students over the six year period. The U_{GPA} and the HS_{AVG} of incoming students in our sample have both remained relatively constant over the period. There has been a slight decrease in the fraction of foreign students and the number of males over the six year period.

[Table 3 here]

---

[5]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 5 years to graduate. Unfortunately, our data set does not capture the respective size of these effects.

8

Regarding the High School variable, we recorded the high school from which the student graduated.[6] In all, there are students from 84 Manitoba high schools in our sample, which are located in 35 Manitoba School Divisions. The high schools are either publicly or privately funded. The privately funded high schools are designated as private independent schools and are included in the Division 8 school division which is classified as Funded (Independent).[7] The majority of schools in are located within the City of Winnipeg.

The school divisions other than Funded Independent Schools are publicly funded institutions in the Province of Manitoba. The students in the publicly funded school divisions in our sample come from 77 public schools, while the Funded Independent School Division includes 7 private schools. During the 1997 - 2002 period, the largest number of first year enrolments came from the Winnipeg School Division (998) followed by Funded Independent Schools (913), Assiniboia (598), Pembina Trails (451), Louis Riel (382) and Seven Oaks (302), and several other School Divisions located in Winnipeg, including St. James. The remaining students came from a large number of rural school divisions. The incoming students are from 84 Manitoba High schools, with the largest number, 316 students, coming from the University of Winnipeg Collegiate, which is in the Funded Independent School Division, and is affiliated with the University of Winnipeg. For the Funded Independent schools we also recorded whether the private school was religious based or secular.

Apart from the school characteristics described above, we recorded several additional characteristics for each student, including their place of residence, the expenditure by the school division

---

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

[7]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.

in which the high school was located and area of study. Unfortunately, the University of Winnipeg does not record the financial background of the students or the educational level of parents or guardians in their admission process. Given that educational attainment is often seen as related to family income, we use as a proxy for a student's family income, the median family income associated with the respective postal code given as their permanent residence.[8] We also collected school expenditures per pupil (for each year and for each school division) from Manitoba Education, Citizen and Youth.[9] Of note, the nominal expenditure per student by school divisions has risen over the period.

In terms of the area of study, we recorded the respective FCEs taken by students in each faculty at the University of Winnipeg. Our sample shows a slight increase in the fraction of FCEs taken in Education, a slight increase in Humanities and a slight decrease in Science FCEs completed over the sample period. Given this information, we then created a "major" dummy variable for each student, defined as the faculty in which they took the largest fraction of their FCEs.[10]

Each observation involves an individual who has a complete record of the characteristics described above.

---

[8]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 neighborhood 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 1st 3 digits of the postal code).

[9]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.

[10]As can be seen this isn't quite like a Major in a subject area, but indicates a concentration of study in a particular faculty. We also used a continuous measure of this variable, the fraction of FCEs taken in a faculty, but the interpretation of this effect is not as straightforward as a dummy variable approach.

3 A Model of University Performance

In order to develop an empirical model of university performance, we follow the Hierarchical Linear Model (HLM) approach and classify our explanatory variables as level 1 or level 2 variables. For our purposes, Level 1 variables describe the student's characteristics, while Level 2 variables describe the characteristics of the high school from which the student graduated. The idea behind the HLM approach is that level 1 units are "nested" or grouped in level 2 units. Our general approach is to estimate the following relationship:

$$U_{GPA} = \ f(\textit{Level 1 variables, Level 2 variables})$$

The dependent variable U_{GPA} , is the University GPA 5 years after the initial enrollment at the University of Winnipeg. The first cohort of students in our sample entered the University of Winnipeg in 1997. In all we recorded the subsequent GPA after five years of students for the entering classes of $1997, 1998, 1999, 2000, 2001$. The minimum GPA recorded at the University of Winnipeg is "1" which is equivalent to "grade D", while the maximum GPA is "4.5", which is equivalent to a grade of "A+". For each entering class, we collected information on the student characteristics and then matched the the student's ID with the University GPA they achieved over the five year period.

The Level 1 variables - the student characteristics - can be defined more specifically as follows. Each year is defined as Summer/Spring, Fall, and the following Winter session. Each Student ID represents a unique student. The variable, High School Average (HS_{AVG}), is calculated as the average of the best marks of 3 approved high school courses (including English, and Mathematics) which is the criteria used for admission to the University of Winnipeg. All students in our sample

attended Manitoba High schools. The variable, AgeDif, is the age of the student upon first entry into the University of Winnipeg minus 18 years. We include a squared term for AgeDif to test for nonlinearity. The variable Family Income, is the Household Income associated with the area (postal code) or permanent resident of the student at the time of admittance. We also include a number of dummy variables as Level 1 variables. The variable Male is 1 if the student is male, 0 otherwise. The variable, Foreign, classifies the students as either Canadian citizens or students from outside Canada. Specifically, the variable Foreign is equal to 1 for international students or students who have landed immigrant status in Canada, and 0 otherwise. We also include as regressors, a set of dummy variables which capture the subject area in which the student took the majority of their courses. The subject areas are Education, Humanities, or Science (with Social Science being the base case). This allows us to test whether the University GPA attained by students varies depending on the particular subject area chosen.

Level 2 variables include the following. The dummy variable Private School is coded 1 if student attended a private school, (which for our data set corresponds to the Funded Independent School Division) and 0 otherwise. The variable Religious is equal to 1 if the student attended a religious based private high school, 0 otherwise. The variable Expend/Student, is the total expenditure per student by the School Division from which the student graduated.[11] A set of time dummies, YR_i indicates the year in which the student first entered the University of Winnipeg.

---

[11]It has been suggested expenditures in different areas by high schools, for example mathematics, may have a greater impact on a student's success in university than other types of expenditures. While data on expenditure per student disaggregated by expenditure category would be desirable, that level of detail is unfortunately unavailable.

3.1 Possible Hypotheses

There are a number of possible hypotheses regarding the expected signs of the regression parameters. Along with much of the literature, we assume that HS_{AVG} is a strong predictor of University performance. We also test for possible nonlinearities regarding the effect of a student'age on the subsequent university performance. The sign of the Foreign student dummy variable, might be negative given the possible academic difficulties faced by students for which English is a second language. Regarding the effect of gender, recent academic research (as well as anecdotal evidence) suggests that females are outperforming their male counterparts in recent years, both at the high school and post secondary level.[12] Regarding the High school effect, it is possible that the subsequent performance of Manitoba high school students is affected by the resources spent on their high school education as well as the non-pecuniary features of the high school (academic standards, discipline). For example, a common perception is that students from private high schools perform better at the University of Winnipeg because of greater high school resources.[13] We attempt to separate these two effects by recording the differences in school expenditures as well as the pure private school effect by identifying whether the student graduated from a public or private school system, and whether the private school was religious based or secular.[14] Financial variables (fam-

---

[12]See for example Cohn et al. (2004) and Davioğlu and Türüt-Aşik (2004) for evidence that females outperform their male counterparts.

[13]In contrast to this result Smith and Naylor (2005) find that the students who attended a Independent school in the U.K. were less likely to obtain a 'good' degree than students who attended a state-sector school.

[14]In a study examining the performance of Finnish senior secondary school students for the years 1990-1998, Häkkinen, Kirjavainen and Uusitalo (2003) find that changes in school spending did not have a signficant effect on test scores. this is consistent with the conclusion of Hanushek that there is no systematic evidence that more school resources improve student learning. Horowitz and Spector (2005) find evidence that students from religious based high schools outperform their private and public school counterparts. Neal (1997) finds that the effect of Catholic secondary schooling differs between groups, with modest educational gains accruing to urban whites, with urban minorities gaining the most.

ily income support, student loans, and scholarships) are also thought to play a significant role in helping students succeed in their post-secondary studies.[15] In particular, the higher is the financial support from the student's family, the higher is student's expected U_{GPA}. However, given that 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.

3.2   The Estimation Procedure

We estimate the the high school fixed effects using two methods. The first approach, given by the Least Squares Dummy Variable estimator (LSDV), estimates the fixed effects by estimating a separate dummy variable for each high school. The second approach, the Hierarchical Linear Model (HLM), assumes the high school effect (or the school intercept) is random. Using the LSDV approach, we correct the standard errors using the Cluster estimator, which takes into account that the error structure is related to the group from which the observation originates. We define the Cluster variable, as the student's High School at graduation. Regarding the second approach, the HLM model is increasingly used in the analysis of education issues, for example, Etherington (1997), and Pike and Saupe (2002). A principal reason is that the students are nested within classrooms and/or schools. HLM models are designed to incorporate this "nesting" feature of the data.

---

[15]There is evidence that the educational attainment of children is strongly effected by the educational attainment of parents, 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.

4   Estimation Results: The Base Model

Prior to presenting the results for the LSDV and HLM estimators, we ran a base model that did not include any school dummies. The estimates for this base model are presented in Table 4. Like Betts and Morrell (1992) we restrict our sample in order to address the issue of a small number of students from a high school having an influential effect on the estimates of their school effects. In our case, we only include those Manitoba high schools that sent at least 3 students to the University of Winnipeg over the sample period. This restriction resulted in a sample size to 5136 observations, and included students from 84 Manitoba High Schools.

Table 4 includes a number of specifications, all estimated with a cluster robust estimator. All specifications are based on the centered independent variable, HS_{AVG}, which is the student's high school average (HS_{AVGi}) minus the grand mean (high school average of all students)in our sample ($\overline{HS_{AVG}}$). The Base model results, Model 1, indicate significant effects of HS_{AVG} (centered around the grand mean), age, choice of major (Education and Humanities) and nationality, and number of FCES chosen (centered around the grand mean of FCEs chosen in our sample). In particular, older students, those majoring in Humanities, and students taking more FCEs, have a higher University GPA after 5 years than younger students majoring in the Social Sciences (the base major).[16] The U_{GPA} of foreign students graduating from Manitoba High schools is estimated to be approximately .21 grade points lower than their Canadian cohorts. We also find a nonlinear effect for the age of students entering the University of Winnipeg, specifically, the U_{GPA} of students increases until age 39 ($21.16 + 18$), then declines after that age.

---

[16]We are aware that the choice of major may be considered as endogenous, which may potentially bias the OLS results. This issue is addressed more fully in section

[Table 4 here]

Models 2 and 3 included private school effects which are designed to test for the effects of private versus public schooling, as well as the effect of private religious instruction. The issue of religious versus secular education has been an intense area of research among educational researchers. Beginning with Coleman et al. (1982, 1987), followed by Goldberger and Cain (1982), the issue of whether Catholic schooling in U.S. provides greater learning outcomes has been a controversial proposition. Much of the discussion relates to the inferences that can be drawn from the data and the accompanying empirical work.

Table 4 addresses the more modest proposal of whether students from public versus private schools, or religious based schools, perform better, on average at the University of Winnipeg, controlling for our measure of family income and school expenditures. The results for Model 2, suggest that students attending a Private School have a U_{GPA} after 5 years that is .138 points higher than students attending a Public School. Model 3 tests whether the religious nature of the private school makes a difference; however, the interaction term is statistically insignificant. Models 4, 5 and 6 test for interaction effects related to the student's High School average, and finds that attending a Private High school increases the effect of high school grades on U_{GPA}, and finds both Private Schooling and School Expenditure increase the effect of high school grades on U_{GPA}. Overall, the R squared for the models range from .44 to .45.

We also examined the performance of students in our sample at various stages of their academic career. Table 4(b) estimates the base model using a student's cumulative U_{GPA} at the end of each year over a five year period. As can be seen the effect of HS_{AVG} as a predictor of U_{GPA} falls in a

16

continuous fashion over the five year period.[17] This is similar to the results of Grove and Wasserman (2004); however, our use of a student's academic performance at the end of each academic year does not indicate a slump in performance. This is in contrast to Grove and Wasserman (2004) who examine student performance on a semester basis. It appears that using an annual time period masks the slump in performance that might occur using observations on a semester by semester basis.

5 LSDV Results

Table 5 lists the results for the LSDV school effects estimator, which includes a separate dummy for each high school. Given that the students are grouped in clusters (in this case, high schools) we use a LSDV(Cluster) estimator. Like Betts and Morrell we find significant differences in the intercepts associated with the individual high schools. The school effects are estimated with a constant term (which compares the respective effects with a base high school).

We also ran the school effects model for different restrictions on the number of students attending the University of Winnipeg from a particular Manitoba high school. The alternative restrictions are that at least 3 students, 10 students or 30 students attended the University of Winnipeg over our sample period. Restricting the sample size to include only the high schools that sent 30 students to the University of Winnipeg, reduces the sample size to 4691 from 5136, with a corresponding reduction in the number of schools from 114 to 38.

[Table 5 here]

---

[17]Table 4(b) indicates 181 fewer observations for the Year 1 regression. This is a result of missing observations for the Year 1 University GPA for these students.

As might be expected, the variation in school fixed effects is larger for larger sample sizes since the larger sample includes some high schools that sent relatively few students to the University of Winnipeg. Restricting the discussion to case of $N_i \geq 30$, the range is .85 University Grade Points. That is, controlling for other determinants of student performance, the difference between the mean U_{GPA} of a student from the lowest performing and highest performing school is almost a full University Grade Point. We also find that neither the financial background of students nor high school expenditures are statistically significant predictors of the student's subsequent GPA at the University of Winnipeg. Regarding the effect of area of study, the higher the percentage of FCEs taken in Humanities and Education raises the students GPA (over the estimate of the Social Science effect) while the higher percentage chosen in Science is negative but not statistically significant.

The entering year time dummies until 2001 are not statistically significant, however, they are negative and significant for 2001 and 2002. That is, controlling for other determinants of university performance, the incoming classes of 2001 and 2002 performed significantly worse than students in previous years. This might be seen as evidence that later students are less well prepared for university than their earlier peers.[18]

6 Hierarchical Linear Model Results

The HLM model was introduced to address the the fact that research data often involves a hierarchical or nested data structure.[19] For example, it is important to note that the data on the

---

[18]In a companion research program, we are investigating the issue of grade inflation, which we feel can be tested using the data set used here.

[19]For a good introduction to the motivation for HLM models, see Raudenbush and Bryk (2002:3-14).

performance of students is "nested", that is, the students are grouped in different classrooms or high schools. In traditional models, the role of the high school is captured by using school dummy variables as in the LSDV approach.[20] While the Least Squares Dummy Variable (LSDV) approach has been extensively used by economists in empirical work, there have been issues raised regarding its use. A major issue is that for estimation problems involving a large number of cross sections, the approach significantly reduces degrees of freedom.

HLM models were developed to provide improved estimation of individual effects for models involving a large number of cross sections.[21] In addition, HLM models can be used to formulate and test hypotheses regarding how variables at one level (student characteristics) are affected by variables are other levels (school characteristics). Finally, HLM models allow the researcher to concentrate on the estimation of variance and covariance components of nested data. That is, it is often important to distinguish between the error variances at different levels, in order to partition the variance of school performance, for example, into within school and between school components.

Following Raudenbush and Bryk (2002) the HLM approach is described as follows. Consider first what has been termed the empty model. That is, suppose we let i denote the index for individuals within the groups ($i = 1, n_j$) and j is the index for the groups ($j = 1, .N$) and let Y_{ij} be the university GPA of student i who graduated from high school j.

---

[20]Within the LSDV approach, it is also possible to test for interaction of the high school dummies with particular slope coefficients, such as β_i.

[21]Raudenbush and Bryk (2002:7-10) nicely summarizes the general use of hierarchical models. For alternative presentations, see Kreft and De Leeuw (1998), Snijders and Bosker (1999), or Rabe-Hesketh, Sophia and Anders Skrondal (2005).

6.1 The Empty Model

The empty model involves no regressors and simply states that the university GPA of student i, graduating from high school j , Y_{ij} is determined by equation (1). The level-1 or student-level model is

$$Y_{ij} = \beta_{0j} + r_{ij} \tag{1}$$

where we assume $r_{ij} \sim$ independently $N(0, \sigma^2)$, where $i = 1, ..n_j$ students in school j, and $j = 1, N$ schools. The parameter σ^2 is the student-level variance. In the empty model, the parameter β_{0j}, which is the mean University GPA for each school, is determined as a function as the grand mean γ_{00} plus a random error, u_{0j} , with the random error assumed to be $u_{0j} \sim$ independently $N(0, \tau_{00})$ where τ_{00} is the school level variance.

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{2}$$

If we substitute for β_{0j} in (1) using (2) we obtain what Raudenbush and Bryk (2002) term the mixed model with fixed effect β_{00} and random effects u_{0j} and r_{ij} which is

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \tag{3}$$

The results from the empty model are listed in the first column of Table 6. What is called the Fixed Effects component in the empty model is the weighted least squares estimate for the grand-mean, $\gamma_{00} = 2.897$ with a standard error of 0.0274, yielding a 95% confidence level of $2.897 \pm 1.96(.0274) = (2.8433, 2.950)$.

The Variance Components is the restricted maximum likelihood estimates of the variance components. At the student level, the estimated variance is, $r_{ij} = 0.63921$ while at the school level

20

the estimated variance of the true school means β_{0j} around the grand mean γ_{00}, which is τ_{00} equals 0.03465. A 95% confidence interval for the school means is $\gamma_{00} \pm 1.96(\tau_{00})^{1/2}$ which equals $2.897 \pm 1.96(.186145) = (2.533, 3.261)$, which indicates a fair range in the performance of the students in the sample.

A useful auxiliary calculated for HLM models is the intraclass correlation, which, in our case, represents the proportion of the variance in University GPA between schools. Using the results for the empty model which are summarized in the first column of Table 5 reveals that the intercept $\gamma_{00} = 2.8973$, which yields an intraclass coefficient of $\rho(Y_{ij}, Y_{i'j}) = \tau_{00}/(\tau_{00} + \hat{\sigma}^2) = .03465/(.03465 + .63924) = .051$. As discussed by Rabe-Hesketh and Skrondal (2005:8) the intraclass correlation "directly measures the 'closeness' of observations on the same subject relative to the closeness of observations on different subjects." As discussed by Snijders and Bosker (1999:46), the intraclass coefficient can either be thought of as "the correlation between two randomly drawn individuals in one randomly drawn group, it is also the fraction of total variability that is due to the group level." A measure often used for evaluating HLM models is the Deviance number which is $-2(LogLikelihood)$ which in this case equals 12367.688746.

6.2 Alternative Specifications of HLM models

The HLM approach allows a wide variety of statistical hypotheses regarding the intercepts and slopes in an econometric model. For example, suppose we introduce a variable to allow us to consider the relationship between the High School Average of students and their subsequent University GPA.

$$Y_{ij} = \beta_{0j} + \beta_{1j}(HS_AVG - \overline{HS_{AVGj}}) + r_{ij} \tag{4}$$

where Y_{ij} is the University GPA in 5 years of student i from high school j, and $HS_{AVG} - \overline{HS_{AVG_j}}$ is the difference between the student i's high school average and the mean high school average at his/her high school. We can define this difference as dHS_{AVGj}, which is described in the HLM literature as a "group centered" variable.

An alternative centering approach that is also used is grand mean centering, which is $dHS_{AVG} = HS_{AVG} - \overline{HS_{AVG}}$ where $\overline{HS_{AVG}}$ is the HS_{AVG} of all students in the sample.

$$Y_{ij} = \beta_{0j} + \beta_{1j}(HS_AVG - \overline{HS_{AVG}}) + r_{ij} \tag{5}$$

In this paper, we use grand mean centering for our HLM model in order to made the results comparable with the LSDV approach.[22] Another reason for grand mean centering is that we are interested in how well high school grades predict the U_{GPA} over all students and not just the relative performance of students within high schools.

The variations of the HLM model that are estimated, essentially involve alternative hypotheses regarding the intercept β_{0j} and slope β_{1j}. Specifically, when β_{0j} and or β_{1j} are assumed to be stochastic, this is called a Random Coefficient Model. When β_{0j} and or β_{1j} are assumed to determined by one or more level 2 variables, the resulting model is called an Intercept and/or Slope as Outcome Model.

Given these preliminaries, it is now possible to specify the complete HLM model used in this paper. In order to reduce the level one variance from the Empty Model, alternative approaches are considered. The approaches generally involve including additional explanatory variables at the student level (Level one). Additional level one variables available in our data set include a number

---

[22]For a discussion of the differences between alternative forms of centering, see Hofmann and Gavin (1998)and Kreft et al. (1995).

of characteristics describing the students, including their age, gender, citizenship (foreign versus Canadian), family income, as well as a series of time dummies indicating the cohort year for each set of students, the entire set of additional level 1 variables is represented by the vector X_{kj}.

Abstracting from the level 2 variables for now, we can rewrite (5) as,

$$Y_{ij} = \beta_{0j} + \beta_{1j}(HS_AVG - \overline{HS_{AVG}}) + \Sigma\beta_{kj}X_{kj} + r_{ij} \tag{6}$$

Note that we in this paper, we assume the coefficients on the set of additional level one variables β_{kj} are non-random.

6.2.1 The Random Coefficient Model

Using (6), Random Coefficient models specify that either β_{0j} the school mean, or β_{1j} the effect of HS_{AVG} on U_{GPA}, could be random. In the following, we allow both the slope and school mean to be random, or

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{7}$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \tag{8}$$

For the case where both β_{0j} the school mean, or β_{1j} are random, we substitute (7) and (8) into (6), which yields

$$Y_{ij} = \gamma_{00} + \gamma_{1j}(HS_{AVG} - \overline{HS_{AVG}}) + u_{0j} + u_{1j}(HS_{AVG} - \overline{HS_{AVG}}) + \Sigma\beta_{kj}X_{kj} + r_{ij} \tag{9}$$

As can be seen the error term is now $u_{0j} + u_{1j}(HS_AVG - \overline{HS_{AVG}}) + r_{ij}$ which as pointed out by Raudenbush and Bryk (2002:21) is not the type of error structure assumed in standard ordinary

least squares estimation. The errors are $u_{0j} + u_{1j}$ are dependent within each school because they are common to every student within school j, and the errors have unequal variances.

In general, the Random Intercept model in the HLM literature is equivalent to the Random Effects (MLE) estimator used in panel data, where the random effects are associated with an identifiable groups or cluster.[23] In our case, the group is the high school (or school division) of graduation. That is, one can interpret the HLM (Random Intercept) results in much the same way as one considers the Random Effects (MLE) results.

6.2.2 The Slopes as Outcomes Model

In the Slopes as Outcomes model, we can allow either the slope and intercept for the school or both to be determined by a level 2 variable, for example, T, school type. We test for two alternative school types; by introducing a dummy variable that indicates whether the student graduated form a private of public high school, and a dummy variable that indicates whether the high school was a religious based or secular school. In the model developed here, we also introduce a second level 2 variable, $Rexp$ which is the real expenditure per student by the school division in which the school resides.[24]

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Rexp_j + \gamma_{02}T_j + u_{0j} \tag{10}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Rexp_j + \gamma_{12}T_j + u_{1j} \tag{11}$$

---

[23]The results using the HLM 6 Hierarchical Linear and Nonlinar MOdeling Software, Raudenbush et al, for the Random Intercept model (without robust standard errors) are equivalent to the Random Effects (MLE) results using Stata. For a further discussion of the equivalence of these results see Rabe-Hesketh and Anders Skrondal (2005).

[24]Unfortunately, data on expenditure per student is not collected at the high school level, but is collected at the school division level. We do not feel this is a significant issue since the school division has the responsibility of ensuring that schools within its boundaries receive approximately equal funding.

Assuming that β_{0j} and β_{1j} determined by (10) and (11), (6) can be rewritten as

$$Y_{ij} = \gamma_{00} + \gamma_{01}Rexp_j + \gamma_{02}T_j + \gamma_{10}(HS_{AVG} - \overline{HS_{AVG}}) + \gamma_{11}Rexp_j(HS_{AVG} - \overline{HS_{AVG}})$$

$$+\gamma_{12}T_j(HS_{AVG} - \overline{HS_{AVG}}) + u_{0j} + u_{1j}((HS_{AVG} - \overline{HS_{AVG}}) + \Sigma\beta_{kj}X_{kj} + r_{ij} \qquad (12)$$

As can be seen, the slopes as outcome approach allows for the level 2 variables to interact with level 1 variables, in this case, $(HS_{AVG} - \overline{HS_{AVG}})$, in addition to directly determining the intercept from (10).[25] Once again, the error structure is similar to that for the random coefficient model.

6.2.3 HLM Results: Summary

Table 6 lists the results from the HLM estimator for the Random Intercept, the combined Random Intercept and Random Slope model and the Slopes as Outcomes case.[26] In terms of the HLM model, column 2 of Table 6 shows that the HLM results are similar to the LSDV estimates with respect to the size and significance of the regressors. In comparison to the Empty model, the magnitude of the level 1 error, as measured by the standard deviation (labelled as R) falls from 0.79951 to 0.59009, with the additional level 1 regressors. The Deviance statistic falls considerably, from 12367.688 for the Empty Model, to 9366.51 for the Random Intercept model and the additional level 1 regressors.[27]

[Table 6 here]

---

[25]It is important to note that the two level 2 variables should not be included in the set of regressors X_{kj} since they are already included as fixed effects in equation (9). If they are, perfect collinearity results.

[26]We used the HLM 6 Hierarchical Linear and Nonlinear Modeling software, by Stephen Raudenbush, Tony Bryk and Richard Congdon, Scientific Software International, Inc. 2000, to estimate the HLM models. The xtmixed estimator from Stata Version 9.2 was used in an earlier version to estimate the HLM models. The results were essentially identical.

[27]The Deviance statistic can be used for model specification, with the change in the Deviance Statistic, being distributed as a chi-squared with $(m_1 - m_o)$ additional parameters estimated, where $m_1 > m_o$. For details see Snijders and Bosker (1999:89).

The results for the Random Slope estimator are given in column 3 of Table 6. The Random Slope estimator allows a coefficient to be a random variable as indicated by (8). In the estimation problem considered here, this means that the relationship between a student's high school average and their University GPA is a random variable. With the addition of a random effect for the slope, β_{1j}, the results reveal little change in either the parameter estimates, or in the Deviance statistic.

In columns (4) and (5) of Table 6, we introduce a level 2 variable, specifically, a dummy variable indicating whether the high school from which the student graduated was public or private. Column 4 assumes that only the intercept is effected by the nature of the High School, which implies, when using (10), the assumption that $\gamma_{01} = 0$. That is, the only level 2 variable effecting the slope β_{1j} is the private school variable. The results from column (4) change little with the inclusion of the private school effect, but it is important to note that the estimate of the private school effect (0.118832 University Grade Points) is significant, but slightly lower than the estimate using the LSDV estimator.

Table 7 summarizes the results from including a number of additional level 2 variables in the HLM model. The variables are the Real Expenditure per student by the student's high school, and whether the private school is religious or secular. Columns 1-3 in Table 7, assume that these variables only effect the intercept, while columns 4-6 assume these variables effect both the intercept β_{0j}, and the slope β_{1j}.

[Table 7 here]

Overall, the size and significance of the level 1 regressors change little for the alternative specifications. The specification with the smallest Deviance statistic, is summarized presented in column

26

(4) of Table 7. The model allows for the two level 2 variables, Private schooling, and Real Expenditure per student, to determine both the intercept β_{0j}, and the slope β_{1j} as given by (10) and (11). The results for the model specified in column (4) of Table 7 suggest the following. First, it appears the relationship between a student's High School Average and University GPA, given by the parameter β_{1j}, is positively related to the Real Expenditure per student and whether the high school was private. This means that there is a stronger relationship between High School Average and University GPA is more spending occurred at a student's high school, and if was a Privately operated school. Regarding the effect of these level 2 variables on the intercept β_{0j}, the Private school effect is not statistically significant, while the Real Expenditure per student lowers slightly the mean grade of students in the sample.

In models that include random effects for both slopes and intercepts, it is customary to examine the statistical relationship between these random variables, in particular, the variance-covariance components. Using the random components from the model specified in Column (4) of Table 6, the estimated variance covariance matrix is

$$\Gamma = \begin{pmatrix} \tau_{0,0} & \tau_{0,1} \\ \tau_{1,0} & \tau_{1,1} \end{pmatrix} = \begin{pmatrix} 0.02394 & 0.00003 \\ 0.00003 & 0.00000 \end{pmatrix}$$

or in terms of correlations

$$\Gamma_1 = \begin{pmatrix} 1.000 & 0.288 \\ 0.288 & 1.000 \end{pmatrix}$$

The off-diagonal of the correlation matrix indicates the correlation between the intercept and slope, given by (10) and (11). The estimated correlation is 0.288, which indicates that the school

27

means and school predictor (HS_{AVG}) are positively correlated. This suggests that high schools from which students subsequently achieve higher mean University GPAs, also exhibit a closer relationship between their mean high school averages and the GPAs achieved at the University of Winnipeg.

7   The Issue of Endogenous Choice of Major

The above models contain a set of regressors describing the main area of study for students. It is clear that there is a potential problem with the estimates, given the fact that the respective choices of majors may be endogenous. In order to test for that possibility, we ran two alternative estimators which are used to address the possible bias created in the structural model by the inclusion of a set of endogenous regressors. The estimator used were an IV estimator and a 2SLS estimator. In both cases, we need to find an exogenous variable which is highly correlated with the endogenous variable but uncorrelated with the error term in the structural equation, but also not a exogenous variable in the structural model.[28]

There are a number of potential IV variables. We examined two possible sets of IV variables, the first, was to include the fraction of the choice of the major chosen by the high school j from which student i graduated. The logic here is that some high schools may have a particular strength in a faculty area, which might influence the choice of major by students. A second set of IV variables was the relative first year GPA of student i, to their overall first year GPA. The idea here is that the higher the relative first year performance of a student in a faculty, the more likely they are to choose that faculty for their major. While it may appear that these variables would be correlated

---

[28]For an up to date discussion of the IV estimation, see Cameron and Trivedi (2005), Chapter 4. See also Angrist and Pischke (2009), Chapter 4 for a nice discussion of IV and 2SLS estimation. For discussion of IV estimation using Stata see Baum (2006), Chapter 8 and Cameron and Trivedi (2009), Chapter 6.

with the error term in the structural equation, the relative scaling, does not distinguish between the absolute performance of students. That is our measure is as likely to explain the choice of major by both relatively strong and relatively weak students.[29]

To test the endogeneity of the student's choice of major, we use as an instrumental variable, a dummy variable which is coded 1, if that major is chosen by the majority of students from student i's high school. We assume that the choice of major by a student is highly influenced, that is, highly correlated with the that chosen by other students at student i's high school, but is uncorrelated with the error term for student i in the structural equation. In the first stage of 2SLS, we ran each student's choice of major against on all exogenous variables and IVs respectively. We, then, saved the predicted values for each student's choice of major respectively. The results for the 1st stage of the 2SLS estimation show that three of the four excluded instruments are highly correlated with student i's choice of major. Based on the Hansen J statistic for over-identification test of all instruments, the results indicate that the null hypotheses that the instruments are uncorrelated with the error term cannot be rejected and suggests which provides evidence in favor of the given specification of the structural equation.

In addition, rather than use the preferred choice of major by other students at student i's high school as an instrumental variable, we also used a dummy variable, which was coded 1, if the student's 1st year GPA from courses taken in a faculty, was the highest among the faculties available at the University of Winnipeg. The 1st stage of the 2SLS results as well as the results for

---

[29]Other IV variables might be the number of University of Winnipeg instructors and professors in the respective faculty, or the respective average class sizes in the faculties. The former suggests the bigger the department, the larger the variety of courses, which may attract relatively more majors. The latter suggests that the smaller the average class size the more likely the student may choose a major in that particular area.

the overidentification test are similar to the results using the alternative IV variables. That is both set of IVs are correlated with student i's choice of major and uncorrelated with the error term for student i in the structural equation. Therefore, we can not reject the consistency of the estimated results using OLS (with robust standard errors), which are still valid.[30]

8 Conclusion

There are a number of conclusions that result from this study. Regarding the LSDV estimates, we find significant high school effects, in particular there is a considerable range in the estimates of the fixed effects from the LSDV estimator. We find that foreign students perform significantly below their peers at the University of Winnipeg despite the fact that all students in our sample graduated from Manitoba high schools. We find that a student's High School average is a strong predictor of their University GPA, but a number of other factors play a significant role in predicting the performance of a Manitoba high school student at the University of Winnipeg. Regarding the HLM estimates, we find, in general a close correspondence with the results from the LSDV estimator.

In future work, we plan to examine the issue of grade inflation, which we feel is a natural extension of the research undertaken here. We feel that our data set allows us to address the issue of grade inflation, in particular, given the fact that all students have graduated from Manitoba high schools, which are governed by Province of Manitoba curriculum.

---

[30]The regression results for both IV estimators are available from the authors on request.

9   Appendix

In this Appendix we provide some information regarding the estimators used in the paper. What follows is based on the description in Raudenbush and Bryk (2002). Aditional details can be found in Chapters 3 and 4 of Raudenbush and Bryk (2002).

In the HLM approach, assume the model is of the form

$$Y_j = X_j \beta_j + r_j \tag{13}$$

where $r_{ij} \sim N(0, \sigma^2 I)$, and Y_j is an n_j by 1 vector of outcomes, X_j is an n_j by $(Q+1)$ matrix of predictor variables, β_j is a $(Q+1)$ by 1 vector of unknown parameters, and I is an n_j by n_j identity matrix, and r_j is an n_j by 1 vector of random errors normally distributed with mean vector of 0 and a variance-covariance matrix in which all diagonal elements are equal to σ^2 and all off-diagonal elements are 0. Given (17) the OLS estimator of β_j is the well known

$$\hat{\beta}_j = (X_j^T X_j)^{-1} X_j^T Y_j \tag{14}$$

with dispersion matrix

$$Var(\hat{\beta}_j) = V_j = \sigma^2 ((X_j^T X_j)^{-1} \tag{15}$$

which yields, premultiplying (17) by $(X_j^T X_j)^{-1} X_j^T$

$$\hat{\beta}_j = \beta_j + e_j \tag{16}$$

where $e_j \sim N(0, V_j)$ and V_j is the error-variance matrix. For HLM models, there is level 2 stage where the general model for β_j is

$$\beta_j = W_j \ \gamma + u_j \tag{17}$$

where $u_j \sim N(0, t)$ where W_j is a $(Q+1)$ by F matrix of predictors, γ is a vector of fixed effects, u_j is a $(Q+1)$ by 1 vector of level 2-errors or random effects, and T is an arbitrary $(Q+1)$ by $(Q+1)$ variance-covariance matrix. If we substitute (21) into (20) we get the combined model

$$\hat{\beta}_j = W_j \ \gamma + u_j + e_j \tag{18}$$

with the dispersion of $\hat{\beta}_j$, given W_j is

$$Var(\hat{\beta}_j) = Var(u_j + e_j) = T + V_j = \Delta_j \tag{19}$$

with the term T described as parameter dispersion and V_j the error dispersion. Given that in general, the groups will have different numbers of observations, the Δ_j will differ from group to group. Assuming that each Δ_j is known, the unique minimum-variance, unbiased estimator of γ is the generalized least squares (GLS) estimator.

$$\hat{\gamma} = (\sum W_j^T \Delta_j^{-1} W_j)^{-1} \sum W_j^T \Delta_j^{-1} \hat{\beta}_j \tag{20}$$

As observed by Raudenbush and Bryk (2002:44) the GLS estimator weights each group's data by its precision matrix, given as Δ_j^{-1}, which is the inverse of the variance-covariance matrix. Given the normality assumptions of the errors r_j and u_j, (24) is also the maximum likelihood estimator of γ.

10 References

Altonji, Joseph G., Todd E. Edler and Christopher R. Taber "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools", *Journal of Political Economy*, 115, 1, 151-184.

Angrist, Joshua D. and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton, N.J.: Princeton University Press, 2009.

Baum, Christopher F., *An Introduction to Modern Econometrics Using Stata*, College Station Texas: Stata Press, 2006.

Betts, Julian R. and Darlene Morell, "The Determinants of Undergraduate Grade Point Average: The Relative Importance of Family Background, High School Resources, and Peer Group Effects", *Journal of Human Resources*, Vol. 34, No.2 (Spring, 1999) pp. 268-293.

Cameron, A. Colin and Pravin K. Trivedi, *Microeconometrics Using Stata*, College Station Texas: Stata Press, 2009.

Cameron, A. Colin and Pravin K. Trivedi, *Microeconometrics: Methods and Applications*, Cambridge: Cambridge University Press, 2005.

Cohn, Elchanan, Sharon Cohn, Donald C. Balch and James Bradley Jr. "Determinants of undergraduate GPAs: SAT scores, high-school GPA and high-school rank" *Economics of Education Review* 23 (2004) 577-586.

Coleman, James S. and Thomas Hoffer, *Public and Private High Schools: The Impact of Communities*, New York: Basic Books, 1987.

Coleman, James S., Thomas Hoffer and Sally Kilgore, *High School Achievement: Public,*

Catholic and Private Schools Compared. New York: Basic Books, 1982.

Ermisch, John and Marco Francesconi, "Family Matters: Impacts of Family Background on Educational Attainments", *Economica* 68, 2001, 137-156.

Feinstein, Leon and James Symons "Attainment in secondary school" *Oxford Economic Papers*, 51, (1999), 300-321.

Goldberger, Arthur S. and Glen C. Cain "The Causal Analysis of Cognitive Outcomes in the Coleman, Hoffer and Kilgour Report" *Sociology of Education*, 55, (1982), pp. 103-22.

Grove, Wayne A. and Tim Wasserman "The Life-Cycle Pattern of Collegiate GPA: Longitudinal Cohort Analysis and Grade Inflation" *Journal of Economic Education*, Spring 2004, 162-174.

Häkkinen, Iida, Tanja Kirjavainen and Roope Uusitalo, "School Resources and student achievement revisited: new evidence from panel data" *Economics of Education Review*, 22 (2003) 329-335.

Hanushek, E. "Assessing the effects of school resources on student performance: an update" *Educational Evaluation and Policy Analysis*, 19, (1997) 141-164.

Hanushek, Eric A. "The Economics of Schooling: Production and Efficiency in Public Schools" *Journal of Economic Literature*, Vol. 24, No. 3 (Sep., 1986), pp. 1141-1177.

Hofmann, David A. and Mark B. Gavin, "Centering Decisions in Hierarchical Linear Models: Implications for Research in Organizations" *Journal of Management*, 24, (1998), 623-641.

Horowitz, John B. and Lee Spector, "Is there a difference between private and public education on college performance?" *Economics of Education Review*, 24, (2005) 189-195.

Kreft, Ita and Jan De Leeuw, *Introducing Multilevel Modelling*, London: Sage Publications, 1998.

Kreft, Ita G. G., Jan de Leeuw, and Leona S. Aiken, "The Effect of Different Forms of Centering in Hierarchical Linear Models, *Multivariate Behavioural Research*, 30(1), (1995), pp. 1-21.

Neal, Derek "The Effects of Catholic Secondary Schooling on Educational Attainment" *Journal of Labour Economics*, 15, 1, (1997) pp. 98-123.

Pike, Gary R. and Joseph L. Saupe, "Does High School Matter, An Analysis of Three Methods of Predicting First-Year Grades", *Research in Higher Education*, Vol. 43, No.2, April 2002, pp. 187-207.

Rabe-Hesketh, Sophia and Anders Skrondal, *Multilevel and Longitudinal Modelling Using Stata*, College Station, Texas: Stata Press. 2005.

Raudenbush, Stephen W. and Anthony S. Bryk, *Hierarchical Linear Models: Applications and Data Analysis Methods* Second Edition, Thousand Oaks: Sage Publications, 2002.

Smith, Jeremy and Robin Naylor, "Schooling effects on subsequent university performance: evidence for the UK university population" *Economics of Education Review*, 24 (2005) 549-562.

Snijders, Tom and Roel Bosker, *Multilevel Analysis: An introduction to basic and advanced multilevel modelling*, London: Sage Publications, 1999.

Zimmerman, David J. "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment" *The Review of Economics and Statistics*, February 2003, 85(1): 9-23.

Table 1: Final Data Set

	# of Missing Obs	# of Obs remaining
Original Dataset		14246
Missing High School Name	5928	8318
Missing High School Marks	1995	6323
Missing School Expenditure	7	6316
Missing Proxy for Family Income	869	5447
Missing Age	1	5446
Missing GPA (5 Yrs)	259	5187
Restricting sample (3 students+ per school at Uof Wpg)	51	5136

Table 2: University GPA, High School GPA and its subcategory breakdown

	OBS.	MEAN	ST.DEV	VAR	MIN	MAX	Median
Student-Level Variables							
U_GPA							
Overall	5136	2.89	0.81	0.66	1	4.5	2.92
Male	1845	2.77	0.83	0.69	1	4.5	2.75
Female	3291	2.96	0.79	0.63	1	4.5	3
Canadian	4967	2.9	0.81	0.65	1	4.5	2.93
Foreign	169	2.56	0.83	0.69	1	4.41	2.46
Public School	4229	2.86	0.8	0.65	1	4.5	2.88
Private School	907	3.02	0.84	0.7	1	4.5	3.03
Graduated (within 5 years)	1814	3.30	0.60	0.36	1.75	4.49	3.29
Did Not Graduate (within 5 years)	3322	2.67	0.82	0.68	1.00	4.50	2.63
HS_AVG							
Overall	5136	78.3	10.03	100.61	51	100	79
Male	1845	76.09	10.44	109.06	51.67	100	75.67
Female	3291	79.54	9.57	91.62	51	100	80.67
Canadian	4967	78.4	10.02	100.36	51	100	79
Foreign	169	75.31	9.97	99.42	52	96.67	74.33
Public School	4229	78.16	9.96	99.25	51.67	100	78.67
Private School	907	78.94	10.32	106.57	51	100	79.67
Graduated (within 5 years)	1814	81.87	9.40	88.39	51.00	100	83
Did Not Graduate (within 5 years)	3322	76.35	9.83	96.56	51.67	99.67	76.67
Age	5136	18.84	2.53	6.38	16	69	18
Family Income/student (2002 in 000's)							
Public School	838	55.15	13.95	195	22.1	101.9	54.4
Private School	190	61.08	20.05	402	22.1	101.9	57.9
School-Level Variables							
Expenditure/student(2002 in 000's)							
Public School	838	6.94	0.49	0.24	5.31	7.71	6.72
Private School	190	6.76	2.02	4.09	4.01	13.38	5.96
School Type	**Number**						
Private	16						
Religious Based Private	(12)						
Public	68						
Total	84						

Table 3: Means of Variables

	1997	1998	1999	2000	2001	2002	Total/ Average
No. of Students (1st year)	754	814	795	830	915	1028	5136
U_GPA	2.84	2.86	2.96	2.88	2.91	2.89	2.89
HS_AVG (%)	77.08	77.24	78.12	78.14	79.19	79.51	78.3
Age (years)	18.75	18.84	18.8	18.98	18.85	18.8	18.84
Male (%)	37.4	37.47	34.47	34.34	35.74	36.19	35.92
Education FCEs (%)	2.86	3.72	4.61	5.45	4.79	4.28	4.31
Humanities FCEs (%)	28.42	28.93	30.03	30.87	30.97	30.5	30.01
Science FCEs (%)	29.76	28.59	28.53	24.83	25.55	25.96	27.08
Social Science (%)	38.96	38.76	36.83	38.85	38.7	39.26	38.59
Total FCEs (per student)	10.78	11.57	11.86	12.04	11.85	11.36	11.58
Foreign (%)	4.11	4.55	2.39	3.73	2.73	2.53	3.29
Expenditure/Student (000's)	5.81	5.98	6.27	6.5	6.65	6.91	6.39
Family Income (000's)	47.18	48.62	51.56	51.2	52.46	56.24	51.49

Table 4: Base Model and School Type Results (Dependent Variable U_GPA)

	(1) Base		(2) Private		(3) Religious		(4) Interaction - Expend		(5) Interaction - Private		(6) Interact Both	
HS_AVG	0.0461***	(0.00140)	0.0459***	(0.00132)	0.0459***	(0.00129)	0.0329***	(0.00724)	0.0442***	(0.00127)	0.0269***	(0.00672)
HS_AVG*Rexp							0.00215	(0.00117)			0.00282**	(0.00102)
HS_AVG*Private									0.00879*	(0.00364)	0.00971**	(0.00318)
AgeDif (Age-18)	0.0707**	(0.0246)	0.0714**	(0.0244)	0.0714**	(0.0246)	0.0713**	(0.0243)	0.0708**	(0.0241)	0.0706**	(0.0240)
AgeDif^2	-0.00167**	(0.000538)	-0.00169**	(0.000539)	-0.00169**	(0.000543)	-0.00169**	(0.000539)	-0.00166**	(0.000528)	-0.00166**	(0.000528)
Male	-0.00870	(0.0197)	-0.0165	(0.0187)	-0.0165	(0.0187)	-0.0174	(0.0188)	-0.0185	(0.0188)	-0.0200	(0.0192)
Education - Major	-0.146***	(0.0333)	-0.136***	(0.0336)	-0.136***	(0.0337)	-0.135***	(0.0336)	-0.134***	(0.0341)	-0.133***	(0.0341)
Humanities-Major	0.101***	(0.0201)	0.104***	(0.0200)	0.104***	(0.0200)	0.104***	(0.0199)	0.107***	(0.0200)	0.108***	(0.0200)
Science-Major	0.0201	(0.0251)	0.0175	(0.0241)	0.0174	(0.0241)	0.0164	(0.0238)	0.0192	(0.0244)	0.0180	(0.0241)
Foreign	-0.212***	(0.0476)	-0.242***	(0.0598)	-0.243***	(0.0610)	-0.246***	(0.0600)	-0.226***	(0.0573)	-0.229***	(0.0568)
Real Expend/student	-0.00175	(0.0168)	0.00149	(0.0166)	0.00132	(0.0165)	-0.00466	(0.0193)	-0.00314	(0.0175)	-0.0117	(0.0195)
Real Family Income	0.00172	(0.00112)	0.00127	(0.001000)	0.00127	(0.00100)	0.00125	(0.00101)	0.00110	(0.000990)	0.00105	(0.000992)
FCEs	0.0320***	(0.00172)	0.0320***	(0.00173)	0.0320***	(0.00173)	0.0320***	(0.00173)	0.0322***	(0.00172)	0.0322***	(0.00172)
1998	-0.0117	(0.0358)	-0.00828	(0.0346)	-0.00846	(0.0344)	-0.00721	(0.0346)	-0.00696	(0.0348)	-0.00542	(0.0348)
1999	0.0354	(0.0362)	0.0348	(0.0357)	0.0345	(0.0357)	0.0364	(0.0359)	0.0365	(0.0359)	0.0388	(0.0361)
2000	-0.0622	(0.0445)	-0.0616	(0.0427)	-0.0618	(0.0425)	-0.0593	(0.0432)	-0.0593	(0.0425)	-0.0561	(0.0434)
2001	-0.0674	(0.0368)	-0.0660	(0.0347)	-0.0662	(0.0344)	-0.0644	(0.0348)	-0.0616	(0.0350)	-0.0591	(0.0355)
2002	-0.0842	(0.0426)	-0.0840*	(0.0415)	-0.0842*	(0.0415)	-0.0815	(0.0416)	-0.0820	(0.0416)	-0.0786	(0.0416)
Private School			0.138***	(0.0358)	0.142***	(0.0327)	0.131***	(0.0360)	0.133***	(0.0384)	0.124**	(0.0384)
Private x Religious					-0.00728	(0.0553)						
Constant	2.793***	(0.110)	2.772***	(0.117)	2.773***	(0.116)	2.810***	(0.132)	2.805***	(0.123)	2.859***	(0.133)
N	5136		5136		5136		5136		5136		5136	
R-sq	0.447		0.451		0.451		0.451		0.453		0.454	

Standard errors in parentheses

="* p<0.05   ** p<0.01   *** p<0.001"

Table 4(b) : Base Case: Regression Coefficients over Student's Academic Career

Dependent variable: University GPA after Year (n)…

Coefficients	Year 1	Year 2	Year 3	Year 4	Year 5
HS_AVG	0.0553*	0.0526*	0.0512*	0.0495*	0.0461*
Agedif (Age-18)	0.0901*	0.0906*	0.0929*	0.0846*	0.0707*
Agedif squared	-0.0020*	-0.0024*	-0.0026*	-0.0022*	-0.0017*
Gendermale	0.0361	0.0194	0.0015	-0.0103	-0.0087
Education[1]	0.0905	-1.8258*	-1.4506*	-0.3046*	-0.1463*
Humanities[1]	0.0874*	0.05**	0.0665**	0.0621**	0.1012*
Science[1]	-0.0131	-0.0239	-0.0135	-0.00003	0.0201
Foreign	-0.1916*	-0.2192*	-0.2217*	-0.2264*	-0.2124*
Real Expenditure	0.0024	-0.0065	-0.0064	-0.0021	-0.0018
Real Income	0.0017	0.0020***	0.0022**	0.0021***	0.0017
Accumulative FCE's[2]	0.0397**	0.0526*	0.0411*	0.0372*	0.0320*
1998	-0.0018	0.0308	0.0085	0.0011	-0.0117
1999	0.0419	0.0886*	0.0558***	0.0596***	0.0354
2000	-0.0538	-0.0195	-0.0294	-0.0293	-0.0622
2001	-0.0732***	-0.0039	-0.0254	-0.0424	-0.0674***
2002	-0.0569	-0.0175	-0.0539	-0.0646	-0.0842***
_cons	2.6275*	2.6227*	2.6612*	2.6944*	2.7925*
N	4955	5136	5136	5136	5136
R-sq	0.3944	0.4516	0.4503	0.4455	0.4469

*: significant at 1% level, **: significant at 5% level, ***: significant at 10% level

Note: [1] Dummy=1 if the completed FCE's in a

particular year is highest in the respective faculty

[2] Student's accumulated FCE's at end of year

Table 5: LSDV Results - School Effects (Dependent Variable U_GPA)

	(1)		(2)		(3)	
	N$_i \geq$3		N$_i \geq$10		N$_i \geq$30	
HS_AVG	0.0470***	(0.00120)	0.0470***	(0.00122)	0.0468***	(0.00127)
AgeDif (Age-18)	0.0884***	(0.0191)	0.0875***	(0.0197)	0.0831***	(0.0207)
AgeDif ^{2}	-0.00188***	(0.000435)	-0.00186***	(0.000444)	-0.00180***	(0.000481)
Male	-0.0262	(0.0188)	-0.0268	(0.0191)	-0.0264	(0.0197)
Education- Major	-0.136***	(0.0331)	-0.133***	(0.0334)	-0.133***	(0.0347)
Humanities-Major	0.0917***	(0.0190)	0.0931***	(0.0192)	0.0890***	(0.0198)
Science-Major	0.0213	(0.0244)	0.0199	(0.0247)	0.0222	(0.0254)
Foreign	-0.255***	(0.0685)	-0.262***	(0.0674)	-0.256***	(0.0693)
Real Expend/ student	0.0378	(0.0459)	0.0371	(0.0474)	0.0346	(0.0472)
Real Family Income	0.00103	(0.000960)	0.000891	(0.000989)	0.000907	(0.00103)
FCEs	0.0314***	(0.00173)	0.0311***	(0.00176)	0.0313***	(0.00183)
1998	-0.0214	(0.0337)	-0.0175	(0.0342)	-0.0252	(0.0360)
1999	0.0161	(0.0361)	0.0138	(0.0368)	0.00365	(0.0384)
2000	-0.0789	(0.0401)	-0.0804	(0.0408)	-0.0786	(0.0424)
2001	-0.0858*	(0.0341)	-0.0840*	(0.0345)	-0.0763*	(0.0357)
2002	-0.113**	(0.0411)	-0.109*	(0.0418)	-0.103*	(0.0433)
Constant	2.762***	(0.256)	2.774***	(0.264)	2.790***	(0.264)
(School Dummies Range)	(-1.522 to .375)		(-0.823 to .371)		(-0.788 to .0632)	
Schools	84		55		38	
Obs	5136		4992		4691	
R-sq	0.482		0.479		0.475	

*= $p<0.05$, **=$p<0.01$ and ***=$p<0.001$

Table 6: HLM Results (Dependant Variable U_GPA)

Final Estimation of Fixed effects (with robust standard errors)										
	Empty Model		Random Intercept		Random Intercept and Random Slope		Intercept as Outcome		Slope and Intercept as Outcomes	
	(1)		(2)		(3)		(4)		(5)	
Fixed Effect	Coefficient.	se	Coefficient	se	Coefficient	se	Coefficient	se		
HS_AVG			**0.04693**	0.001201	**0.046569**	0.001422	**0.046891**	0.001197		
Intercept 2									**0.044919**	0.001266
Private School									**0.010951**	0.003756
AgeDif (Age-18)			**0.085570**	0.000449	**0.085154**	0.019836	**0.085657**	0.019942	**0.084656**	0.019527
AgeDif ^{2}			**-0.001855**	0.001631	**-0.001839**	0.000445	**-0.001858**	0.00045	**-0.001822**	0.000435
Male			-0.021310	0.018021	-0.022190	0.018080	-0.02195	0.017992	-0.024742	0.017842
Foreign			**-0.249693**	0.062621	**-0.241745**	0.061863	**-0.252237**	0.063185	**-0.238239**	0.060785
Education-Major			**-0.133058**	0.032365	**-0.129506**	0.032632	**-0.132399**	0.032457	**-0.127989**	0.032931
Humanities-Major			**0.093933**	0.018741	**0.095856**	0.018835	**0.094762**	0.01877	**0.097929**	0.019007
Science-Major			0.018606	0.024003	0.018034	0.024211	0.01817	0.02397	0.018874	0.02432
Real Family Income			0.001212	0.000936	0.001117	0.000915	0.001137	0.000922	0.000946	0.000906
FCEj			**0.031446**	0.001691	**0.031609**	0.001695	**0.031465**	0.001693	**0.031695**	0.001689
1998			-0.018939	0.033183	-0.016836	0.033121	-0.018016	0.033035	-0.015098	0.033195
1999			0.022513	0.034245	0.022460	0.034184	0.0228	0.034173	0.023515	0.034448
2000			-0.068015	0.040386	-0.066167	0.040338	-0.067486	0.040353	-0.066095	0.040115
2001			**-0.077398**	0.033342	**-0.074274**	0.033516	**-0.07712**	0.033275	**-0.072947**	0.033554
2002			**-0.098873**	0.038678	**-0.096094**	0.038585	**-0.098202**	0.038573	**-0.096607**	0.038779
Intercept1	**2.8970**	-0.0274	**2.770128**	0.053815	2.771582	0.053630	2.75069	0.055372	**2.757943**	0.055663
Private School							**0.118832**	0.047277	**0.096884**	0.049112

Final Estimation of variance components										
Random Effect	Standard Deviation	Variance Comp	Standard Deviation	Variance Comp	Standard Deviation	Variance Comp	Standard Deviation	Variance Comp	Standard Deviation	Variance Comp
Intercept 1, U0	0.18616	0.03465	0.16172	0.02615	0.15841	0.02509	0.15124	0.02287	0.15387	0.02368
HS_AVG slope					0.00417	0.00002			0.00274	0.00001
Level - 1, R	0.79951	0.63921	0.59009	0.3482	0.58899	0.34691	0.59026	0.34841	0.58856	0.34641
Deviance	12367.68875		9366.513762		9361.44511		9364.244696		9356.208472	
Observations	5136		5136		5136		5136		5136	

Table 7: HLM Results (II)　　　　　　　(Dependant Variable U_GPA)

Final Estimation of Fixed effects (with robust standard errors)

	Intercept as Outcome		Intercept as Outcome		Intercept as Outcome		Intercepts and Slopes as Outcomes		Intercepts and Slopes as Outcomes		Intercepts and Slopes as Outcomes	
	(1)		(2)		(3)		(4)		(5)		(6)	
Fixed Effect	Coefficient	se	Coefficient	se	Coefficient	se	Coefficient	se	Coefficient	se	Coefficient	se
HS_AVG, B3												
Intercept 2	**0.046907**	0.001193	**0.046895**	0.001199	**0.046910**	0.001191	**0.045023**	0.001195	**0.044989**	0.001210	**0.045033**	0.001192
Rexpend							**0.003407**	0.000893			**0.003214**	0.000796
Private School							**0.012713**	0.002836	**0.006709**	0.002628	**0.010753**	0.001589
Private x Rel									0.006234	0.005490	0.003484	0.004285
AgeDif (Age-18)	**0.085738**	0.019878	**0.085731**	0.019944	**0.085769**	0.019856	**0.084437**	0.019494	**0.084578**	0.019639	**0.084305**	0.019511
AgeDif ^{2}	**-0.001859**	0.000449	-0.001859	0.000450	**-0.001858**	0.000449	**-0.001817**	0.000435	**-0.001817**	0.000438	**-0.001811**	0.000436
Male	-0.021817	0.017976	-0.021983	0.018013	-0.021934	0.018001	-0.025145	0.017985	-0.024297	0.017863	-0.025113	0.017971
Foreign	**-0.252294**	0.063349	**-0.252250**	0.063245	**-0.252440**	0.063612	**-0.243189**	0.060873	**-0.240786**	0.062347	**-0.245309**	0.061330
Education-Major	**-0.133051**	0.032415	**-0.132436**	0.032474	**-0.132976**	0.032439	**-0.126354**	0.032904	**-0.128691**	0.032952	**-0.126595**	0.032973
Humanities-Major	**0.094892**	0.018749	**0.094721**	0.018754	**0.094809**	0.018744	**0.098311**	0.019008	**0.097186**	0.018829	**0.097760**	0.018930
Science-Major	0.018536	0.023937	0.018214	0.023935	0.018491	0.023917	0.017754	0.023895	0.018952	0.024097	0.017660	0.023852
Real Fam Income	0.001143	0.000922	0.001135	0.000922	0.001139	0.000923	0.000917	0.000909	0.000942	0.000904	0.000917	0.000910
FCEj	0.031454	0.001693	**0.031462**	0.001693	**0.031445**	0.001693	**0.031670**	0.001689	**0.031632**	0.001700	**0.031631**	0.001697
1998	-0.017803	0.033002	-0.018021	0.032989	-0.017995	0.032990	-0.014592	0.033227	-0.014983	0.033232	-0.014133	0.033300
1999	0.023605	0.034104	0.022799	0.034146	0.023460	0.034151	0.024685	0.034621	0.023150	0.034546	0.024407	0.034695
2000	-0.066641	0.040335	-0.067456	0.040260	-0.066896	0.040293	-0.065481	0.040418	-0.065751	0.040068	-0.065193	0.040414
2001	**-0.076375**	0.033199	**-0.077087**	0.033178	**-0.076672**	0.033193	**-0.072925**	0.033569	**-0.073708**	0.033430	**-0.073170**	0.033557
2002	**-0.097923**	0.038573	**-0.098232**	0.038549	**-0.098158**	0.038569	**-0.095863**	0.038969	**-0.096987**	0.038692	**-0.095863**	0.038925
Intercept1, B0												
Intercept 2	2.747979	0.055170	2.750605	0.055339	**2.747902**	0.055026	**2.757013**	0.055404	**2.758659**	0.055462	**2.757136**	0.055391
Rexpend	-0.017849	0.015802			-0.022097	0.017328	**-0.034589**	0.017608			-0.034509	0.019348
Private School	**0.123924**	0.051847	**0.112870**	0.050815	**0.169996**	0.060323	0.093105	0.054816	0.099765	0.061081	0.114249	0.061161
Private x Rel			0.007249	0.075350	-0.060509	0.079456			-0.002998	0.084408	-0.029715	0.087654

Final Estimation of variance components

Random Effect	Standard Deviation	Variance Comp.	Standard Deviation	Variance Comp	Standard Deviation	Variance Comp	Standard Deviation	Variance Comp	Standard Deviation	Variance Comp	Standard Deviation	Variance Comp
Intercept 1, U0	0.152400	0.02323	0.15334	0.02351	0.15495	0.02401	0.15472	0.02394	0.15617	0.02439	0.15748	0.02480
HS_AVG slope							0.00076	0.00000	0.00141	0.00000	0.00079	0.00000
Level - 1, R	0.590220	0.34836	0.59025	0.34839	0.59017	0.34831	0.58837	0.34618	0.58883	0.34672	0.58835	0.34615
Deviance	9371.293862		9368.589646		9371.502609		9361.960007		9365.85724		9372.588902	
Observations	5136		5136		5136		5136		5136		5136	