DISCUSSION PAPER SERIES

IZA

IZA DP No. 3098

Do College-Bound High School Students Need an Extra Year? Evidence from Ontario's 'Double Cohort'

Louis-Philippe Morin

October 2007

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Do College-Bound High School Students Need an Extra Year? Evidence from Ontario's 'Double Cohort'

Louis-Philippe Morin
University of Ottawa
and IZA

Discussion Paper No. 3098
October 2007

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3098
October 2007

ABSTRACT

Do College-Bound High School Students Need an Extra Year? Evidence from Ontario's 'Double Cohort'*

The Local Average Treatment Effect (LATE) interpretation of the IV estimates of the returns to schooling is becoming increasingly popular. Typically, researchers reporting LATE estimates do not provide systematic evidence that there is substantial heterogeneity across different ability levels in returns, and without such evidence, the LATE interpretation is short of being compelling. The recent abolition of Grade 13 in Ontario's secondary school system provides a unique opportunity to measure the benefits of an extra year of high school for high-ability students (those bound for college), rather than dropouts. I present a simple factor model which allows the value-added of Grade 13 (in terms of achievement) to be estimated, generalizing the standard difference-in-differences estimator to correct for heterogeneity in ability measurement across college subjects. The main finding is that the estimated return to an extra year of high school in terms of human capital is small for these high-ability students: students coming out of Grade 13 have a 2.2 point advantage (on a 100 point scale) over students from Grade 12, the estimated return to Grade 13 being around 2 percent. This evidence indicates that there is substantial heterogeneity in the return to an additional year of high school in the direction assumed in the prior literature.

JEL Classification: I20, I21, I28

Keywords: return to schooling, factor model, difference-in-differences

Corresponding author:

Louis-Philippe Morin
Department of Economics
University of Ottawa
55, Laurier Avenue East
Ottawa, Ontario, K1N 6N5
Canada
E-mail: lmorin@uottawa.ca

---

* The views, opinions, findings, and conclusions expressed in this paper are strictly those of the author. For confidentiality reasons, the data used in this paper cannot be released by the author. All requests about these data should be directed to the Faculty of Arts and Science at the University of Toronto.
I thank Aloysius Siow, Michael Baker, and Robert McMillan for their invaluable guidance. Thanks also to Victor Aguirregabiria, Joshua Angrist, Pierre Brochu, Mark Evans, Xavier Fazio, Phil Oreopoulos, and Silvana Simone for helpful comments. I am grateful to George Altmeyer, Khuong Doan, Susan Pfeiffer, Pekka Sinervo, and the Faculty of Arts and Science at the University of Toronto for making this project possible. All remaining errors are mine.

1 Introduction

Since Angrist and Krueger's landmark 1991 study of the returns to an additional year of schooling, a number of economists have found surprisingly high rates of return to education. For example, Card (2001) lists eight credible studies using features of the school system as instruments, finding returns in excess of 10 percent; these IV estimates are higher than or equal to the corresponding OLS estimates.[1] Such findings have led economists – following Imbens and Angrist (1994) – to interpret the IV estimates as giving the Local Average Treatment Effect (LATE): in general, if there is heterogeneity within a population in the response to a treatment, then the IV estimate will only capture the response from individuals who were affected by the instrument. Lang (1993) and Card (1995) provide an economic justification for the IV-estimate LATE interpretation in the context of estimating the return to schooling: the instruments used in the literature could affect portions of the population with a higher-than-average return to schooling (higher-than-average discount rates), resulting in finding IV estimates higher than OLS estimates (higher than the population average return to schooling). This phenomenon has been labelled 'discount rate bias' by Lang (1993).[2]

Since the instruments used in the literature mainly affect potential school leavers (dropouts), high returns to education from the IV literature are often linked to students with low educational attainment.[3] If there were no heterogeneity in the response to the treatment, then the LATE justification for differences in OLS and IV estimates of the return to schooling would appear questionable (in this case, any LATE should give us the Average Treatment Effect). In particular, the Lang and Card models would not explain these differences.

Although the LATE interpretation of the IV estimates of the return to education is increasingly popular, researchers reporting LATE estimates typically do not provide systematic evidence of substantial heterogeneity in the returns to schooling. Some studies (e.g. Aavik, Salvanes and Vaage

[1]Since ability is usually omitted from standard Mincer's (1974) earnings equations and since ability is expected to be positively correlated with earnings and schooling, estimated returns to schooling from OLS regressions (in the absence of any other problems) should be upward-biased. If the IV estimate represents the population average return to education and the only problem with the OLS estimation is omitted variables, then we would expect OLS estimates to be higher than the IV estimates.

[2]A necessary condition for this phenomenon is that the return to schooling declines as schooling increases. See Angrist and Krueger (1999) and Card (1999) for discussions of discount rate bias.

[3]For example, Angrist and Krueger (1991), Harmon and Walker (1995), Staiger and Stock (1997), Meghir and Palme (2005), and Oreopoulos (2006) use either reforms or variables (e.g. quarter of birth) affecting the minimum legal number of years of schooling of individuals as an instrument.

2003) find evidence of heterogeneity in the returns to schooling across different levels of education but these pieces of evidence do not suggest heterogeneity *within* a given level of education.[4] Without such evidence, Lang and Card's LATE interpretation is plausible but short of being compelling.

In this paper, I investigate the impact of an extra year of schooling on a group of students which, despite its size, has attracted little attention in the literature. Most studies concerned with estimating the return to secondary education concentrate on low educational attainment, representing the left tail of the academic ability distribution. Here, I focus on students who will attend university – the right tail of the ability distribution.[5,6] Since high school serves mainly as college preparation for college-bound students, I concentrate on college academic performance as means of capturing the value-added associated with an extra year of high school.[7]

A recent reform of the Ontario Secondary School (OSS) curriculum provides a unique opportunity to capture the benefit of an extra year of schooling for high-ability students, allowing me to analyze the value-added of the final year of high school for students who will attend university. Motivated by a desire to conform with a majority of North American secondary school curricula and by the possibility of lowering costs in the educational system, the Ontario government announced in 1997 that it would compress its secondary school curriculum. Thus, starting in 1999, students were expected to graduate from high school after four years (i.e. after Grade 12) instead of five (after Grade 13). In 2003, as a consequence of the abolition of Grade 13, two different groups of students graduated from Ontario high schools and entered university simultaneously so that Ontario universities had students with either four or five years of high school in their classrooms at the same time, starting in September 2003.[8]

If the reform could be thought of as a random experiment, then simply comparing students' university performance would capture the value-added of an extra year of high school. But important features of the reform prevent us from measuring the return to Grade 13 using such an

[4] Meghir and Palme (2005) present results for which a sample was divided into high- and low-ability students. They do not find significant differences in the return to schooling between high- and low-ability for male students.

[5] Note that a majority (more than 55%) of Ontario students will now graduate from either college or university. In contrast, the dropout rate among 20 year-olds in Ontario is about 10%.

[6] One paper which also focuses on university-bound students is Krashinsky (2006). Krashinsky (2006) and this present paper both look at the impact of the same Ontario Secondary School reform on students' university performance. An important difference between the two papers is that students studied in Krashinsky (2006) seem to have lower academic ability than students in this study.

[7] Since the outcome variable studied in this paper is academic performance and not earnings, I will refer to the benefit of schooling as "the value-added of" and not "the return to" schooling in order to avoid any potential confusion.

[8] For this reason, the cohort of 2003 high school graduates was known as the Ontario "Double Cohort."

estimation strategy. First, it is likely that the reform had behavioural effects which could lead to serious estimation biases if not taken into account. Students from the double cohort expected that university admission standards would increase dramatically when the two cohorts of students graduated from high school. It is possible that some students who were expected to graduate from Grade 13 in 2003 'fast-tracked' high school and graduated a year early in order to avoid having to face the increased competition for university places. As a result, these students would be missing when comparing the performance of Grade 12 students to Grade 13 students in the double cohort. If students who succeeded in 'fast-tracking' were better-than-average students, a university performance comparison of Grade 12 and Grade 13 students would lead to a biased estimator of the value-added of Grade 13. In particular, we could incorrectly fail to reject the null hypothesis of zero value-added of Grade 13.

Second, the 1997 reform did not affect subjects in a uniform manner. The compression of the secondary school curriculum clearly affected the delivery of material for some subjects and not others. This heterogeneity in treatment across subjects allows me to circumvent the problem associated with comparing two potentially different student populations. By observing students' university performance in at least two subjects (for example, mathematics versus biology), one of which was not affected by the reform, it becomes possible to control for potential differences in ability across groups and to get identification of the value-added of Grade 13.

Third, if ability were measured in the same way across subjects, one could use standard difference-in-differences techniques to capture the value-added of Grade 13. This is likely to be an unrealistic assumption. To account for this problem, I present a simple and flexible factor model which takes into account the possibility that students might differ in ability across groups but also the possibility that subjects do not measure ability in the same way. In the end, the identified value-added from the factor model is a generalization of the standard difference-in-differences estimator which corrects for heterogeneity in ability measurement across subjects.[9] The model also makes it possible to test for other potentially important effects of the reform, such as the presence of high school grade inflation.[10]

[9] An implicit assumption made in this paper is that the amount of time spent studying each subject was not affected by the reform. This is a possible caveat. However, a study by King *et al.* (2004), which looks at the evolution of double cohort students in high school, shows that a majority of university-bound students spend less than 11 hours a week on homework. This suggests that the study time constraint is not binding.

[10] The model is estimated using a flexible GMM estimation. This way I can relax the normality assumption of the

The main finding of the paper is that, for these high ability students, the estimated (human capital) return to an extra year of high school is small: students coming out of Grade 13 have a 2.2 point advantage (on a 100 point scale) over students from Grade 12. To put this in perspective, the standard deviation, in points, is about 13.[11] We can convert this value-added of Grade 13 into a LATE estimate of the return to Grade 13. According to the literature, a one-point increase in GPA translates into a 9-10 percent increase in earnings;[12] using these estimates, the return to Grade 13 would be around 2.2 percent, which is smaller in order of magnitude than the LATE and OLS estimates of the returns to schooling from the previous literature. In particular, this estimate is far below the 6-12 percent found in Krashinsky (2006) who looks at the impact of the same reform on university-bound students with lower high school averages than students studied in this paper.[13] As such, the estimated LATE of this study provides solid evidence that there is substantial heterogeneity in the return to an additional year of high school in the direction suggested by Lang and Card. Further, it underlines the importance of the LATE interpretation in the estimation of the return to schooling.

This paper is organized as follows: I present characteristics of the 1997 Ontario Secondary School Reform which allow for the identification of the value-added of Grade 13 in the next section. Data are described in Section 3. Section 4 presents results from estimating the value-added of Grade 13 using popular estimation methods such as simple mean comparison, difference-in-differences and OLS regression. Shortcomings of these methods in the context of the OSS Reform are then discussed briefly. A model which accounts for these shortcomings is introduced in Section 5, and parameters identification and estimation strategies are presented in sections 6 and 7. Key results are reported in Section 8. Finally, robustness of the results is discussed in Section 9 while Section 10 concludes.

observed variables, which is rejected by the data.

[11] The effect size is 0.17σ. Krashinsky (2006) finds effect sizes above 1.2σ. See Angrist and Lavy (1999), Krueger (1999), and Hoxby (2000) for dicussions on effect size in the context of class size reduction.

[12] See Jones and Jackson (1990) and Loury and Garman (1995).

[13] Comparing my results to Krashinky's (2006) is certainly more sensible than comparing to studies found in Card (2001) since we both look at the effects of the same reform on different student populations, but both comparisons suggest significant heterogeneity in the returns to schooling.

2 The Ontario Secondary School Reform

This section highlights features of the Ontario Secondary School (OSS) curriculum reform that will help in identifying the value-added of Grade 13.

In 1997, the provincial government of Ontario announced that it would compress its secondary school curriculum from five to four years. This reform would bring Ontario into line with most surrounding provinces and potentially lower the costs of the educational system in a significant way. Thus, starting in 1999, students were expected to graduate from high school after four years (after Grade 12) instead of five. In 2003, the first cohort of students from the new curriculum graduated from high school, and in the same year, Grade 13 was also abolished. Thus, in 2003, Ontario colleges had students with two different high school backgrounds in the same classes: some students had four years of high school (henceforth referred to as 'G12' students), while others had five ('G13' students).

If assignment to G12 or G13 were random, then performance of these two groups should capture the benefits of an extra year of schooling. In practice, selection issues might arise, making the impact of the reform harder to identify. Fortunately, other characteristics of the reform will help us to overcome selection issues, beginning with the intensity of the treatment effect.

The intensity of the treatment effect on university preparation should not be seen as being uniform across subjects: the reform did not simply force students to take one less year of schooling. Even though students were now expected to graduate after four years instead of five, they still had to complete the same number of credits (30) as their predecessors in order to satisfy the OSS Diploma requirements.[14] We might think that students from the two curricula (G12 and G13) learned the same material. But college-bound students – who now represent a majority of students – also need to satisfy college admission requirements, which depend on the program they plan to attend.[15]

A quick inspection of changes in two subject-specific high school curricula (biology and mathematics) illustrates the heterogeneity across subjects in the effects of the reform on the amount of material taught to university-bound students. Figure 1 illustrates the transition between the old

[14] The Ontario Ministry of Education and Training (1999) defines a credit as "a means of recognition of the successful completion of a course for which a minimum of 110 hours has been scheduled."

[15] In a recent study presented by King *et al.* (2002), more than 80% of Grade 11 students interviewed were planning to attend college (31.1%) or university (50.5%).

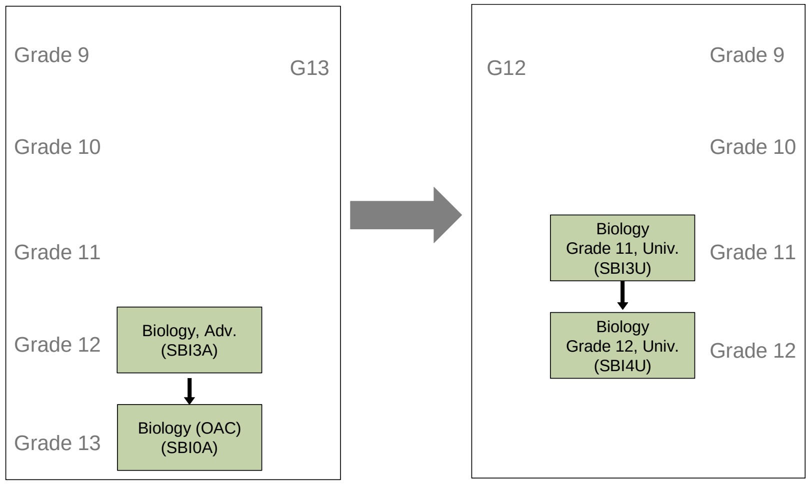

Figure 1: Reform of the Biology Sequence for University-Bound Students

and new biology course sequences imposed on typical university-bound life-science students.[16] Prior to taking a biology course, both groups should have successfully completed Grade 9 and Grade 10 Science courses. Despite the reform, the amount of biology material taught in high school is similar for both groups. G12 students have to take essentially the same two courses which were offered in the G13 program.[17]

While the impacts of the reform on biology and on a majority of subjects were minimal, this is not true for mathematics and the English course sequences. For these subjects, obtaining the senior high school year credit requires a sequence of prerequisites starting in Grade 9. Figure 2 illustrates the transition from the G13 to G12 curriculum of the mathematics sequence followed by a typical university-bound high school student. The reform clearly affected the sequence of courses. Under the new system, students are now expected to take four courses of mathematics instead of five. The amount of material covered in class was affected: less material was covered, and probably less time was spent on each topic.[18]

If the reform caused behavioural responses from students of the form discussed in the Intro-

[16] All students interested in pursuing a life-science university education should complete a sequence of two biology courses prior to attending university. This was true for students enrolled in the G13 curriculum and it is still true today for G12 students.

[17] Comparison of covered-topics description of these two biology sequences confirms the similarity between the two sequences. See Ontario Ministry of Education (1987, 2000b).

[18] Comparison of covered-topics description of these two mathematics sequences suggests that some material which used to be covered in the late stage of the G13 sequence (e.g. integration) tended not to be covered in the G12 sequence. See Ontario Ministry of Education (1985, 2000a).

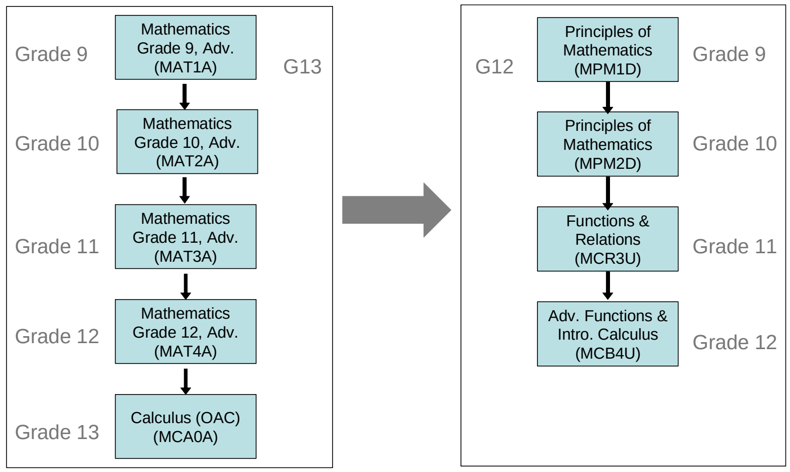

Figure 2: Reform of the Mathematics Sequence for University-Bound Students

duction, then the experiment can no longer be assumed to be random and identification of the value-added, by comparing grades in one subject, is impossible. The heterogeneity in the treatment intensity will be very useful in identifying the value-added of Grade 13 in the presence of selection issues. By observing students' university performance in at least two subjects – biology and mathematics – one of which was not affected by the reform, it becomes possible to control for potential unobserved differences across groups and to get identification of the value-added of Grade 13.[19]

3 Data

The student data used in this study are provided by the Faculty of Arts and Science of the University of Toronto, one of the largest universities in North America. These administrative data contain information about students' first-year university academic performance (e.g. grades, dropped courses, program[20]), and pre-admission academic history (e.g. high school average, identification of secondary school institutions attended, and an indicator of secondary school curriculum graduated from – G12/G13). The data also contain each student's date of birth, gender, and her/his student

[19] Even if the biology curriculum was not affected by the reform, the amount of biology-specific human capital acquired by the students could be affected if study-time constraints were affected. However, variation in time inputs on homework does not seem to have been important. This supports the hypothesis that the amount of biology-specific human capital acquired by the students was not affected by the reform.

[20] In 2003, students interested in studying at the University of Toronto Faculty of Arts and Science had to apply to one of the following programs: Commerce, Computer Science, Humanities and Social Sciences, and Life Sciences.

number.

One advantage of using administrative data for this type of study is that the observations are practically error-free. For example, since we have an indicator of the secondary school curriculum attended by each student (G12/G13 indicator), we do not have to rely on her date of birth to decide which curriculum the student graduated from, though the date of birth allows us to concentrate on the population we are most interested in, namely students born in 1984 and 1985.

In order to capture the value-added of Grade 13, I analyze the effects of Grade 13 on mathematics performance. If there really is a positive effect of Grade 13 on university preparation, we would expect to see the strongest effects on mathematics performance since it is one of the subjects that was most affected by the reform.

I restrict the sample to students enrolled in the Life Sciences program. The advantages of doing so are numerous. First, this is a large program which allows the researcher to observe students taking both a course affected by the reform – mathematics – and another which was not – biology. Second, these subjects are likely to be "independent" in that knowledge of biology should not affect a student's knowledge of mathematics and vice versa.[21]

The third advantage of focusing on Life Sciences is that students interested in a Life Sciences discipline have to complete a list of compulsory courses during their first year of university. This allows me to alleviate course selection issues. All first year students must take the same biology course (BIO150Y), and almost all programs require an introductory calculus course (MAT135Y). About 90% of students for whom we observe a grade for BIO150Y also had a grade for MAT135Y.[22]

Finally, Life-Sciences students' backgrounds, except for G12/G13 differences, are similar. Before joining the Life Sciences program, G12 students must have successfully completed Advanced Functions and Introductory Calculus while G13 must have Calculus. These two courses are the standard university-preparation courses of their respective curricula. Students should also have a senior high school biology credit.[23] Hence, students are expected to have completed both course sequences of their respective curriculum; these are shown in Figures 1 and 2.

[21] English was not analyzed in this paper for this reason. I could not find a program in which we observe students taking both English and another subject "independent" of it.

[22] This is true for both groups of students. It is not surprising to observe such a high proportion of students taking mathematics as well as biology since students may be uncertain about their exact preferences in terms of field of specialization and might simply insure against this uncertainty.

[23] Most fields in the Life Sciences (39 out of 43) require students to have a senior high school biology credit.

Table 1: Descriptive Statistics

G12 (N=502)	Mean	Std. Dev.	Min	Max	Mean Diff. (G13-G12)
Age	18.2	0.3	17.8	18.7	(-)
Female	0.64	0.5	0	1	(-)
HS Average	90.8	3.4	83.0	98.8	(-)
Number of Courses	5.8	0.6	3	8	(-)
G13 (N=436)	Mean	Std. Dev.	Min	Max	Mean Diff. (G13-G12)
Age	19.2	0.3	18.8	19.7	1.0
Female	0.67	0.5	0	1	0.03
HS Average	90.9	3.2	83.7	99.2	0.01
Number of Courses	5.7	0.5	4	7	-0.1

Table 1 presents descriptive statistics on these students. Aside from the age difference, the two groups of students seem very similar: they take the same number of university courses and are both composed of a majority of female students with excellent high school averages.[24] These students seem to have higher academic ability than students from Krashinsky (2006). Students studied in Krashinsky (2006) have a high school average around 84 percent while students studied in this paper have a 91 percent high school average. This difference is considerable: the 'average' student found in Krashinsky (2006) has a high school average close to the minimum average found in this study (83 percent) and about two standard deviations below this group's average.

4 Estimating the Value-Added of Grade 13

Consider the situation where two factors influence a student's average mathematics performance when comparing G12 and G13 – the curriculum taken and student ability. The expected difference in mathematics performance (Δ_M) could then be characterized by the sum of the value-added of G13 (Δ_V) and the difference in average initial level of ability between G12 and G13 (Δ_η).[25] Thus,

$$\Delta_M = \Delta_V + \Delta_\eta. \tag{1}$$

If the OSS reform could be thought of as a random experiment, we might expect the difference in the average level of (initial) ability to be negligible ($\Delta_\eta \simeq 0$). Then the difference in mathematics

[24] Twelve students were excluded from the original sample since they had grades below 30%. See Section 9.
[25] The initial level of ability is the general level of academic acquired prior to secondary schooling.

performance would fully capture the effect of the reform:

$$\Delta_M = \Delta_V.$$

Table 2 presents the students' performance in MAT135Y. Students from the G12 curriculum do not do much worse than students with one more year of high school. In fact, we cannot reject the hypothesis that both groups perform at the same level. G13 students do better than G12 students

Table 2: University Mathematics Performance

Mathematics	Mean	Standard Error
G13 (N=436)	70.61	0.62
G12 (N=502)	70.16	0.59
G13-G12$\equiv \Delta_M$	0.45	0.86

which is what we would expect. The results nevertheless suggest that the value-added of Grade 13 is very small, if not zero, since the difference in performance is not statistically significant. If, however, the randomness assumption of the experiment is violated, then our results are no longer valid. In particular, if the two groups have different levels of ability, then the effect of Grade 13 could be diluted by the difference in ability.

Notice that in equation (1), if both Δ_V and Δ_η are different from zero, there is no way to disentangle the value-added from the difference in ability. In particular, if G12 students have a higher average level of ability than G13 students, then G12 students' ability could compensate for lack of knowledge usually acquired in Grade 13.

4.1 The Ontario Double Cohort

This subsection discusses how the randomness assumption of the double cohort experiment is violated and why the value-added captured by simple mean comparison confounds the value-added of Grade 13 with the difference in average ability between Grade 12 and Grade 13 students.

Since two cohorts of students were expected to graduate from secondary school simultaneously in June 2003, the double cohort created an expected surge of applicants for post-secondary institutions for September 2003. We can see the dramatic increase in the number of Ontario university applicants clearly in Figure 3. Between 2001 and 2003, the number of applicants (per year) in-

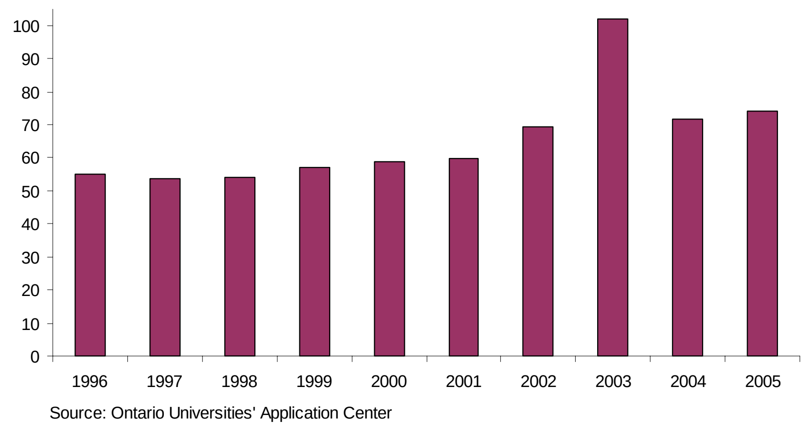

Figure 3: Number of Ontario University Applicants (in thousands)

creased from about 60,000 to close to 102,000. This increase was expected both by students and parents since the announcement of the reform (1997), and is likely to have given rise to behavioural effects.

The expected increase in the number of applicants for 2003 led naturally to expectations of increased competition for university admission. This led some students to try to avoid the double cohort. For example, it was possible under the G13 curriculum to "fast-track" the program and graduate after four years,[26] with the fear of the double cohort probably encouraging some G13 students to try to fast-track and graduate in 2002 instead of 2003.

This idea is supported by Figure 3. The increase in the number of applicants between 2001 and 2002 in the figure is important. The number of applicants rose by about 16% (from 60,000 to 69,000), which is much larger than the average increase prior to 2001, suggesting that some G13 students successfully escaped from the double cohort.[27] One could argue that mainly "high" ability students were able to escape in this way. If high ability G13 students disappeared from the 2003 cohort, the average ability of 2003 G13 students would probably be lower than the average ability of 2003 G12 students. Also, because of the double cohort, some G12 students were encouraged to take five years to complete secondary school or to take some time off between secondary school and university. Even if it might seem costly for a student to delay her university application, it

[26]Even though it was possible to fast-track secondary school, this was far from being common practice. Prior to 2002, around 8% of Ontario university students had graduated from high school after four years.

[27]Demographics cannot explain such increase. The number of 19 year-olds in Ontario increased by 3.4% in 2002.

could be beneficial if she thought that there was a high probability that she would not be accepted into her intended program due to the increased competition in admissions in 2003. If we think that this behavior is more likely to occur among "low" ability students, then we have even more reason to think that the groups of students present in 2003 are different not only in terms of the curriculum they took but also in terms of their average ability level, in which case the estimator of the value-added of Grade 13 would be biased downward when comparing mathematics performance.

Matching students' university grades to their high school average, hoping to control for ability differences across groups, would not alleviate the problem caused by the double cohort. This is because the increased competition for university admission might have also led to grade inflation.[28,29] The model presented in Section 5 will explicitly take this potential problem into account.

4.2 Difference-in-Differences Estimation of the Value-Added

Observing more than one university outcome for each student can help to control for the difference in ability. We could use biology as a proxy for ability and regress the university mathematics grade on the university biology grade and a dummy variable equal to 1 for G13 students and 0 otherwise:

$$M_i = \alpha + G13 * I_i + \beta B_i + \epsilon_i$$

where M_i is the student's university mathematics grade, α is a constant, I_i is an indicator variable equal to 1 if the student is from the G13 curriculum and B_i is the student's university biology grade. $G13$ measures the value-added of Grade 13.

Table 3 presents OLS regression results, suggesting that the value-added of Grade 13 is small but significant. They also suggest that the two groups are different since the estimated value of Grade 13 is close to four times the estimated value when comparing the mathematics performance through a simple means comparison, as in Table 2 (1.68 vs 0.45). A potential problem with the OLS interpretation is that we assume that biology measures ability perfectly. If not, the measure

[28] In this paper, grade inflation means that one group has been graded more (or less) severely than the other. One type of grade inflation that cannot be identified here arises if both groups had their grades increased by a same amount.

[29] G12 students might have been treated favorably relative to G13 by high school teachers to compensate for the stress from being the first cohort of the new secondary school curriculum. Alternatively, G13 that finished high school might have been favored relative to G12. This could be due to the absence of high ability G13 students from the 2003 cohort which could have made it easier to get good high school grades for the G13 students of this cohort.

Table 3: OLS Regression Results

Dependent variable: University Math. Grade		
Independent variables	Coefficient	Standard Error
Univ. Bio. Grade	0.84	0.03
$G13$	1.68	0.63
α	6.40	2.25
R^2	0.47	
N	938	

of the value-added of Grade 13 will be biased.[30]

Since the biology sequence is assumed not to have been affected by the reform, we can use the difference in average biology grades (Δ_B) as a measure of the difference in ability (Δ_η)

$$\Delta_B = \Delta_\eta \tag{2}$$

We can then construct a difference-in-differences estimator using equations (2) and (1):

$$\begin{aligned} \Delta_{DD} &\equiv \Delta_M - \Delta_B \\ &= \Delta_V + \Delta_\eta - \Delta_\eta \\ &= \Delta_V \end{aligned} \tag{3}$$

The difference between differences in average university mathematics grades (Δ_M) and in average biology grades (Δ_B) would give us the value-added of Grade 13 (Δ_V). Table 4 presents the difference in biology performance that we use to control for ability differences in the standard difference-in-differences estimator. A t-test suggests that G12 students do significantly better in biology than G13 students, which also suggests that we are facing two different groups in terms of ability levels.

Table 4: University Biology Performance

Biology	Mean	Standard Errors
G13 (n=436)	74.31	0.51
G12 (n=502)	75.79	0.48
G13-G12$\equiv \Delta_B$	-1.48	0.70

Table 5 presents the difference-in-differences estimate of the G13 value-added based on equa-

[30]The sign of the bias will depend on the groups' relative performance in university biology. See Appendix A.2 for details.

tion (3). The difference-in-differences estimate is more than four times greater than the average difference in mathematics performance. It is also statistically significant. The estimate is precisely estimated but it is still small when compared to the students' mathematics average (70.4) and standard deviation (13.1).

Table 5: Difference-in-Differences Estimator

Math - Bio	Mean	Standard Errors
G13 (n=436)	-3.70	0.48
G12 (n=502)	-5.62	0.42
G13-G12$\equiv \Delta_{DD}$	1.92	0.64

The average G12 student would have had a 2.7% increase in her mathematics performance in the absence of the reform. For many students, this difference would not affect their GPA. Only students close a grade cut-off (e.g. between an A and a B) might see their GPA suffer from missing Grade 13.[31]

The difference-in-differences estimator is only valid if one of the following two assumptions holds: 1) if students' average ability levels do not differ across groups or 2) if biology and mathematics measure students' ability in exactly the same way. To see that this is true, assume that biology does not measure ability in the same way that mathematics does, so that

$$\Delta_B = \lambda^B \Delta_\eta \tag{4}$$

where $\lambda^B \neq 1$. Then

$$\begin{aligned} \Delta_{DD} &= \Delta_M - \lambda^B \Delta_\eta &(5)\\ &= \Delta_V + \frac{(1-\lambda^B)}{\lambda^B}\Delta_B &(6) \end{aligned}$$

where equation (6) is obtained using equations (1) and (4). If both assumptions fail ($\Delta_B \neq 0$ and $\lambda^B \neq 1$) then the difference-in-differences estimator will be biased. I have already shown that we should be suspicious about the former assumption: I now show that we should also be wary of the latter.

Intuitively, if mathematics and biology measure ability the same way, they should have the

[31]Grade B covers scores from 70 to 79. Grade A covers covers scores 80 and above.

same relationship with a third measure of ability. For example, we should expect students' biology and mathematics grades to have the same covariance with their overall high school average. But when we look at Table 6, we can see that the sample covariances between the high school average and biology and mathematics differ. The difference is consistent across groups. The covariance

Table 6: Means and Covariances of Students' Grades

G13 (n=436)	Mean	High School	Biology	Mathematics
High School	90.86	10.1		
	(0.2)	(0.6)		
Biology	74.31	16.7	114.1	
	(0.5)	(1.6)	(7.6)	
Mathematics	70.61	20.5	91.5	167.8
	(0.6)	(2.1)	(7.5)	(11.3)
G12 (n=502)		High School	Biology	Mathematics
High School	90.79	11.8		
	(0.2)	(0.6)		
Biology	75.79	20.6	115.3	
	(0.5)	(1.7)	(6.3)	
Mathematics	70.16	24.3	101.0	176.3
	(0.6)	(2.1)	(6.9)	(10.1)

Note: Covariances are presented in the last three columns.
Standard errors are in parentheses.

between biology and high school is between 15 and 20 percent smaller than the covariance between mathematics and high school (e.g. 16.7/20.5). Not only might the two groups differ in ability, but the two measures of ability used to capture the value-added of G13 might not do so in the same way.

Since we know that Δ_B is negative, the sign of the bias will depend on whether $\lambda^B > 1$ or $\lambda^B < 1$. If $\lambda^B < 1$, the difference-in-differences will be downward-biased, which could explain why we have such a small estimate of the value-added of G13.

This section presented potential problems linked to using simple intuitive techniques (means comparison, OLS estimation and difference-in-difference) when estimating the value-added of Grade 13. The next section presents a model which will be used to circumvent these important potential problems. The model, despite its apparent complexity, will prove to be easily linked to the previous estimation techniques.

5 A Grading Policy Model

I model the relationship between human capital accumulation and academic performance in the specific environment of the double cohort. In particular, the model shows how available observables – university grades, high school grades, and high school curricula – can be linked. It is constructed such that the estimator of the G13 value-added is a generalization of the standard difference-in-differences estimator presented above.

Using information from a third measure of ability – a student's high school average – the estimator from grading policy model is able to account for differences in ability measurement between mathematics and biology and incorporate these in the estimator. The model allows for grade inflation at the high school level, different levels of average ability across groups, and heterogeneity of ability measurement across subjects.

The first step in constructing the model is to define the concepts involved in identifying the effect of Grade 13 on student performance. This section introduces these concepts in the context of a simple grading policy model before moving on to elaborate the model to be estimated.

5.1 Factors Influencing Student Academic Performance

In simple terms, a student's grade can be thought of as the product of three factors: the student's *academic ability*, the school *grading policy*, and a *curriculum effect*.

Students begin high school with an initial level of general *academic ability*. Instead of seeing this ability as solely defined by the individual's innate characteristics or innate ability, we will view it as the joint product of the individual's own innate, acquired, and environmental characteristics.[32] For example, genetics, acquired study habits, and family resources could represent three of these attributes. Notice that they should all be considered to be student-specific characteristics which combine to produce the student's own general *academic ability*. Academic ability is assumed to be partially unobservable; neither the econometrician nor teachers can perfectly measure it.[33]

A *grading policy* is a tool that teachers and professors use to signal (via a grade) a student's subject-specific human capital. Different subjects measure this human capital differently. If we think that performance in each subject measures some aspect of general academic ability, we could

[32] See Cunha, Heckman, Lochner and Masterov (2005) for a discussion of life-cycle skill formation.

[33] All the econometrician observes are grades.

think that they might measure different types of skill, in which case there could be different grading policies for different courses. I assume that the grading policy is under the control of the teacher or the professor.

Third, student performance is likely to be influenced by a group-specific *curriculum effect* when compared to students with different school curriculum backgrounds. In the context of the OSS reform, the difference in the curriculum effect between G13 and G12 students represents the value-added of G13. The curriculum determines how much human capital a group will acquire during high school. Its effect is not only group-specific but also subject-specific. This reflects the fact that a curriculum change can affect some subjects more than others.

5.2 Model

I now present the model more formally. In order to do so, I first need to model the way that human capital is accumulated through high school and university and then model how this human capital is signaled by teachers at each of these levels.

5.2.1 Human Capital Accumulation

There are two institutions superscripted by uppercase I in the model through which students accumulate subject-specific human capital: high school ($I = H$) and university ($I = U$). Assume there are only two subjects $S = \{B, M\}$[34] that a student takes in both high school and university. Two groups of students take different curricula $C = \{G12, G13\}$ while in high school, and are then reunited in university. Hence, in general, model coefficients will have three different superscripts.

Student i is initially endowed with a level of general academic ability (η_i) and then accumulates subject-specific human capital as she attends high school and university. While in high school, $G13$ students receive a treatment which affects the amount of mathematics-specific human capital they acquire. $G12$ students do not receive the treatment, and high school biology is not affected by the treatment. Hence, both groups are assumed to acquire the same biology material in high school. By the end of students' first year of university, they will have accumulated both biology-specific and mathematics-specific human capital (respectively $\eta_i^{U,B}$ and $\eta_i^{U,M,C}$) discussed below.

[34] B stands for Biology while M stands for Mathematics.

Equations (7) and (8) describe the human capital accumulation processes which will drive students' academic performance in university (denoted by the U superscripts on both left hand side variables):

$$\eta_i^{U,B} = \eta_i + \tau^{H,B} + \tau^{U,B} \quad (7)$$

$$\eta_i^{U,M,C} = \eta_i + \tau^{H,M,C} + \tau^{U,M} \quad (8)$$

The first term on the right-hand side of each equation is the student's initial level of general academic ability, the second term is the amount of subject-specific human capital acquired through high school (superscripted by H), and the third term is the amount of subject-specific human capital acquired during first year of university (superscripted by U).

The amount of subject-specific human capital accumulated in high school and university ($\tau^{H,B}$, $\tau^{U,B}$, $\tau^{H,M,C}$, and $\tau^{U,M}$) should be seen as functions of the amount of material taught and the time spent on the material.[35] Since high school biology was not affected by the treatment, both groups of students are assumed to acquire the same amount of biology-specific human capital in high school, so I can suppress the curriculum superscript ($\tau^{H,B,G13} = \tau^{H,B,G12} = \tau^{H,B}$). G12 and G13 students also acquire the same amount of human capital in university ($\tau^{U,B}$ and $\tau^{U,M}$) since they are in the same classes. In contrast, the amount of mathematics-specific human capital accumulated in high school will depend on the curriculum attended by the student. Thus we might have $\tau^{H,M,G13} \neq \tau^{H,M,G12}$. For this reason, the amount of mathematics-specific human capital accumulated by the end of the first year of university ($\eta_i^{U,M,C}$) will depend on the curriculum taken by student i.

5.2.2 Grading Policies

For each student, we observe three grades: 1) a mathematics university grade (M), 2) a biology university grade (B) and, 3) an overall high school average (H). These grades signal the student's relative subject-specific levels of human capital when compared to her classmates. Since the two groups are separated prior to university, high school grades only represent performance with respect to the student's own group (i.e. G12 or G13). For example, in high school biology, a teacher would

[35] This model does not disentangle the effects of 'time spent on material' from the effect of 'more material'. One could imagine $\tau = f(time, material)$, where both time and material have a positive effect on $\tau^{H,M,C}$.

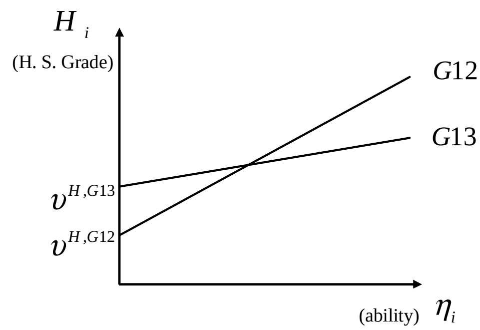

Figure 4: Possible High School Grading Policies

compare a student's level of human capital $(\eta_i + \tau^{H,B})$ to the group average $(E_C[\eta_i] + \tau^{H,B} \equiv \bar{\eta}^C + \tau^{H,B})$. Since students from the same group learn the same material, a student's grade will only depend on her initial general academic ability (η_i) compared to the group average $(\bar{\eta}^C)$.

The high school grading policies (for biology and mathematics) are assumed to be linear in the student's relative level of human capital $(\eta_i - \bar{\eta}^C)$. As a consequence, the high school average (H_i^C) is also a linear function of the difference between the student's initial academic ability and the average initial ability of the group she belongs to:

$$H_i^C = \pi^{H,C} + \lambda^{H,C}(\eta_i - \bar{\eta}^C) + \varepsilon_i^{H,C}$$

The slope and the intercept coefficients (respectively $\lambda^{H,C}$ and $\pi^{H,C}$) represent averages of slope and intercept coefficients across high school subject grading policies. These coefficients are under the teacher's control. We can rewrite the high school average more simply as

$$H_i^C = \upsilon^{H,C} + \lambda^{H,C}\eta_i + \varepsilon_i^{H,C} \tag{9}$$

where

$$\upsilon^{H,C} = \pi^{H,C} - \lambda^{H,C}\bar{\eta}^C.$$

Figure 4 illustrates possible high school grading policies for G12 and G13 students. Both the intercept and the slope coefficients are allowed to vary across high school curricula. The difference in $\upsilon^{H,G13}$ and $\upsilon^{H,G12}$ represents grade inflation. Grade inflation has two potential sources: one source comes from the way teachers link ability to grades, and is measured by $\pi^{H,C}$ and $\lambda^{H,C}$. The other source comes from the student population within each group and is captured by $\lambda^{H,C}\bar{\eta}^{C}$. This term captures the notion that the better are the students from a group, the harder it is to achieve a high grade within this group. The slope coefficient ($\lambda^{H,C}$) represents the payoff to ability. If $\lambda^{H,C}=0$, then high school grades are distributed randomly across students and academic ability plays no role. The error term ($\varepsilon_i^{H,C}$) represents shocks due to measurement error and possible shocks to student performance (e.g. bad luck or illness). $\varepsilon_i^{H,C}$ is assumed to have mean 0 and is uncorrelated with the student's ability. Notice that only the left-hand-side variable of equation (9) is observed.

In the same fashion, the university biology grading policy is given by

$$B_i^C = \pi^B + \lambda^B(\eta_i^{U,B} - \bar{\eta}^{U,B}) + \varepsilon_i^{B,C} \tag{10}$$

Students are now compared to classmates from both groups. For this reason, we have $\bar{\eta}^{U,B}$ and not $\bar{\eta}^{U,B,C}$ in the grading policy equation. I assume that professors do not discriminate against students based on their high school background.[36] The constant and the slope coefficients are then assumed to be the same for both groups. Similar to the high school grading policy, the error terms represent shocks that can be due to simple measurement error but also to temporary shocks affecting students' performance. These error terms are assumed to be uncorrelated with a student's ability but also uncorrelated with each other ($E(\varepsilon_i^{B,C}, \varepsilon_i^{H,C})=0$). We can rewrite (10) more simply as a function of the student's initial level of general academic ability and the population average initial level of general academic ability

$$B_i^C = \upsilon^B + \lambda^B\eta_i + \varepsilon_i^{B,C} \tag{11}$$

where

$$\upsilon^B = \pi^B - \lambda^B\bar{\eta}.$$

$\bar{\eta}$ represents the total population initial level of general academic ability. The difference between the average biology performance ($\bar{B}^{G13} - \bar{B}^{G12} = \Delta_B$) is given by equation (4), i.e. $\Delta_B = \lambda^B\Delta_\eta$.

[36] Typically, university professors do not know the high school background of individual students.

The mathematics high school sequence was affected by the reform. Thus, we can imagine that both the student's initial level of human capital and her curriculum will affect her grade, yielding:

$$M_i^C = \pi^M + \lambda^M(\eta_i^{U,M,C} - \bar{\eta}^{U,M}) + \varepsilon_i^{M,C}$$

or

$$M_i^C = \upsilon^{M,C} + \lambda^M \eta_i + \varepsilon_i^{M,C} \tag{12}$$

where

$$\upsilon^{M,C} = \pi^M - \lambda^M \bar{\eta} + \lambda^M \left(\tau^{H,M,C} - \frac{N_{G13}}{N} \tau^{H,M,G13} - \frac{N_{G12}}{N} \tau^{H,M,G12} \right)$$

with N_{G13} being the number of G13 students and N is the total number of students ($N = N_{G13} + N_{G12}$). The last term of the constant, in parentheses, represents the effect of the curriculum on the students' performance. The difference between $\upsilon^{M,G13}$ and $\upsilon^{M,G12}$ represents the value-added of Grade 13.

The grading policy model consists of equations (9), (11) and (12). We can easily see the resemblance to a standard one-factor model where the driving factor is the initial level of general academic ability (η_i). One necessary condition for identification of the factor model parameters is that the latent variable (η_i) must be scaled to one observed variable. That is, the slope and the intercept coefficients of one equation should be predetermined. Usually, the choice of the benchmark is irrelevant, but since I am interested in the difference between $\upsilon^{M,G13}$ and $\upsilon^{M,G12}$, a convenient normalization is to set the constant and slope coefficient of the mathematics grading policy for G12 students ($\upsilon^{M,G12}$ and λ^M) equal to 0 and 1 respectively. This normalization implies that mathematics professors compensate students such that the direct grade inflation (the grade inflation under their control) cancels out the indirect grade inflation (due to the population average ability). Notice that, after the normalization, $E(M_i^{G13} - M_i^{G12}) = \Delta_V + \Delta_\eta$, just as in equation (1).

Overall, the grading policy model can be summarized in six equations:

$$H_i^{G13} = \upsilon^{H,G13} + \lambda^{H,G13}\eta_i + \varepsilon_i^{H,G13} \tag{1.9a}$$

$$H_i^{G12} = \upsilon^{H,G12} + \lambda^{H,G12}\eta_i + \varepsilon_i^{H,G12} \tag{1.9b}$$

$$B_i^{G13} = \upsilon^{B} + \lambda^{B}\eta_i + \varepsilon_i^{B,G13} \tag{1.11a}$$

$$B_i^{G12} = \upsilon^{B} + \lambda^{B}\eta_i + \varepsilon_i^{B,G12} \tag{1.11b}$$

$$M_i^{G13} = \Delta_V + \eta_i + \varepsilon_i^{M,G13} \tag{1.12a}$$

$$M_i^{G12} = \eta_i + \varepsilon_i^{M,G12} \tag{1.12b}$$

where only the left-hand sides of each equation are observable.

6 Identification

The grading policy model is summarized by a system of equations in which the correlation between the observables (the left-hand-side variables) is due to a single common factor (η_i). In fact, the only difference between this model and a pure factor model is that I allow the equations to have constant terms. The identification strategy follows the approach used in the factor models literature closely.[37] In order to identify the G13 value-added, I will use the basic hypothesis of these models which stipulates that, if the model is correct, the covariance matrix (Σ^C) of curriculum C's observed grades should be exactly reproduced by the covariance matrix implied by the model ($\Sigma(\Theta)^C$), so that

$$\Sigma^C = \Sigma(\Theta)^C \tag{13}$$

where

$$\Sigma^C = \begin{bmatrix} var(H_i^C) & cov(H_i^C, B_i^C) & cov(H_i^C, M_i^C) \\ cov(H_i^C, B_i^C) & var(B_i^C) & cov(B_i^C, M_i^C) \\ cov(H_i^C, M_i^C) & cov(B_i^C, M_i^C) & var(M_i^C) \end{bmatrix}$$

[37] See Hayduk (1987) or Bollen (1989) for details.

and

$$\Sigma(\Theta)^C = \begin{bmatrix} [\lambda^{H,C}]^2\sigma^2_{\eta^C} + \sigma^2_{\varepsilon^{H,C}} & \lambda^{H,C}\lambda^B\sigma^2_{\eta^C} & \lambda^{H,C}\sigma^2_{\eta^C} \\ \lambda^{H,C}\lambda^B\sigma^2_{\eta^C} & \left[\lambda^B\right]^2\sigma^2_{\eta^C} + \sigma^2_{\varepsilon^{B,C}} & \lambda^B\sigma^2_{\eta^C} \\ \lambda^{H,C}\sigma^2_{\eta^C} & \lambda^B\sigma^2_{\eta^C} & \sigma^2_{\eta^C} + \sigma^2_{\varepsilon^{M,C}} \end{bmatrix}$$

where $\sigma^2_{\eta^C} = var(\eta^C)$ and $\sigma^2_{\varepsilon^{B,C}} = var(\varepsilon^{B,C})$. We should expect to have the same kind of relation between the observed-grades' first moments, μ^C, and the first moments implied by the model, $\mu(\Theta)^C$. Hence:

$$\mu^C = \mu(\Theta)^C \tag{14}$$

where

$$\mu^C = \begin{bmatrix} E(H_i^C) \\ E(B_i^C) \\ E(M_i^C) \end{bmatrix} \quad \text{and} \quad \mu(\Theta)^C = \begin{bmatrix} \upsilon^{H,C} + \lambda^{H,C}(E_C\left[\eta_i\right]) \\ \upsilon^B + \lambda^B(E_C\left[\eta_i\right]) \\ I_{G13}\Delta_V + (E_C\left[\eta_i\right]) \end{bmatrix},$$

where I_{G13} is an indicator variable equal to 1 if the student is a G13 student and $E_C\left[\eta_i\right]$ is the average level of initial general academic ability of group C. The model has a total of 18 measured moments and there are 17 coefficients to be estimated (Δ_V, $E_{G13}\left[\eta_i\right]$, $E_{G12}\left[\eta_i\right]$, $\sigma^2_{\eta^{G13}}$, $\sigma^2_{\eta^{G12}}$, υ^B, λ^B, $\upsilon^{H,G13}$, $\upsilon^{H,G12}$, $\lambda^{H,G13}$, $\lambda^{H,G12}$, $\sigma^2_{\varepsilon^{H,G13}}$, $\sigma^2_{\varepsilon^{H,G12}}$, $\sigma^2_{\varepsilon^{B,G13}}$, $\sigma^2_{\varepsilon^{B,G12}}$, $\sigma^2_{\varepsilon^{M,G13}}$, and $\sigma^2_{\varepsilon^{M,G12}}$).[38] The plausible 'no-discrimination' (based on high school curriculum) assumption about the university grading policies and the normalization of the G12 university mathematics grading policy allow for the identification of the model parameters. The simplicity of the model makes it easy to write the parameters of interest as functions of population moments.[39] The slope coefficient of the biology grading policy is

$$\lambda^B = \frac{cov(H_i^C, B_i^C)}{cov(H_i^C, M_i^C)}. \tag{15}$$

Looking ahead, a testable restriction of the model is that the ratio of covariances in equation (15) should be the same for both groups of students.

The value-added of G13 is

$$\Delta_V = \Delta_M - \frac{cov(H_i^C, M_i^C)}{cov(H_i^C, B_i^C)}\Delta_B \tag{16}$$

[38]Each group has six second-moments and three first-moments.

[39]For simplicity, I only present explicit expressions of the main coefficients of interest. See Appendix A.1 for more details about the identification of the parameters.

where $\Delta_M = E(M_i^{G13}) - E(M_i^{G12})$ and $\Delta_B = E(B_i^{G13}) - E(B_i^{G12})$. Equation (16) is equation (5) written differently since the ratio of the covariances is simply λ^B. The estimator of the value-added is thus a modified difference-in-differences estimator allowing for different measures of ability across courses.

We can link (16) to the naive estimators presented above. If biology and mathematics were to measure ability in exactly the same way, then equation (16) would become the standard difference-in-differences estimator. If the two groups of students were identical, then equation (16) would become a simple mean comparison. Finally, if students' high school and biology grades were identical ($H_i^C = B_i^C$), then equation (16) would give the OLS estimator.

The groups' average initial ability levels ($E_C(\eta_i)$) are defined by:

$$E_{G12}(\eta_i) = E(M_i^{G12}) \tag{17}$$

$$E_{G13}(\eta_i) = \frac{cov(H_i^C, M_i^C)}{cov(H_i^C, B_i^C)}\Delta_B + E(M_i^{G12}). \tag{18}$$

From equations (17) and (18), we can see that the student performance in the university biology course will sign the difference in average ability ($\Delta_\eta \equiv E_{G13}[\eta_i] - E_{G12}[\eta_i]$) since $1/\lambda^B$ is positive.

7 Estimation

One could use instrumental-variables techniques to get consistent estimates of these parameters, estimating the model equation by equation. I adopt a different approach, estimating the whole system simultaneously using GMM.[40] I choose GMM over maximum likelihood (ML) since the normality assumption required for the validity of the ML is rejected for all six outcomes. Nevertheless, the results are not sensitive to the choice of estimation strategy.[41]

The empirical strategy is to fit the sample moments to the moments implied by the model. For each group, we have a fit function defined by

$$F(\boldsymbol{\theta})_C = (\mathbf{s}^C - \boldsymbol{\sigma}(\boldsymbol{\theta})^C)' \mathbf{W}_C^{-1}(\mathbf{s}^C - \boldsymbol{\sigma}(\boldsymbol{\theta})^C) + (\mathbf{x}^C - \boldsymbol{\mu}(\boldsymbol{\theta})^C)' \mathbf{S}_C^{-1}(\mathbf{x}^C - \boldsymbol{\mu}(\boldsymbol{\theta})^C) \tag{19}$$

[40]The simplicity of the model makes the results from the two estimation strategies similar. The conclusions were the same whether I used IV or GMM.

[41]Results from the maximum likelihood estimation are available upon request.

where $\mathbf{x}^C$ and $\mathbf{s}^C$ are vectors of sample first and second moments respectively while $\boldsymbol{\mu}(\boldsymbol{\theta})^C$ and $\boldsymbol{\sigma}(\boldsymbol{\theta})^C$ are vectors of first and second moments implied by the model. $\boldsymbol{\theta}$ is the vector of parameters I wish to estimate, $\mathbf{S}_C$ is the sample covariance matrix and $\mathbf{W}_C^{-1}$ is a weight matrix to be defined.

The first part of the equation is the standard GMM fit function used in the analysis-of-covariance literature.[42] The second part is the fit function for first moments which is necessary to estimate the coefficient of G13's value-added. The results are obtained using as the weight matrix $\mathbf{W}_C^{-1}$ estimates of the fourth-order moments. This estimator is the Optimal Minimal Distance (OMD) estimator. Following concerns about the use of the OMD estimator voiced by Altonji and Segal (1994), I also used different weight matrices $\mathbf{W}_C^{-1}$ to check for any disparities in the parameters estimates due to the choice of the weight matrix. The use of the OMD weight matrix, the identity matrix, or a diagonal weight matrix using fourth-order moments as weight matrices all give very similar results.[43]

The global fit function used in the minimization problem is a weighted average of the groups' fit functions

$$F(\boldsymbol{\theta}) = \frac{N_{G13}}{N} F(\boldsymbol{\theta})_{G13} + \frac{N_{G12}}{N} F(\boldsymbol{\theta})_{G12}$$

and the parameter estimates are given by $\boldsymbol{\theta}_{OMD}{=}ArgMin_{\theta} F(\boldsymbol{\theta})$. These are discussed in the following section.

8 Results

Table 7 presents results of the GMM estimation using the OMD estimator weight matrix. The value-added of G13 is positive and precisely estimated. Controlling for ability, Grade 13 increases a student's mathematics performance by 2.2 percentage points. Comparing the value of G13 to the mathematics average and standard deviation,[44] the return to Grade 13 is modest in terms of human capital accumulation. We can get a ballpark LATE estimate of the return to G13 using existing literature. Loury and Garman (1995) and Jones and Jackson (1990), for example, found that a

[42]See Abowd and Card (1987, 1989), Altonji and Segal (1996), Baker (1997), or Altonji, Martins and Siow (2002) for applications of this estimation technique in the labour economics literature.

[43]These results are available upon request.

[44]See Table 6.

one-point in GPA would lead to about 9-10 percent earnings increase.[45] Using these estimates, and assuming that the 2.2 points would translate directly to the GPA, the return to Grade 13 would be around 2.2 percent.[46]

Table 7: Parameter Estimates (GMM Estimation)

	G13	G12
υ (intercept)		
MATH	2.21	(-)
	(0.09)	(-)
BIO	17.21	17.21
	(2.76)	(2.76)
HS	78.23	76.55
	(1.15)	(0.99)
λ (slope)		
MATH	1.00	1.00
	(-)	(-)
BIO	0.83	0.83
	(0.04)	(0.04)
HS	0.18	0.20
	(0.02)	(0.01)
$\bar{\eta}$ (ability)	68.39	70.16
	(0.08)	(0.01)

All three subject grading policies are different. The difference between the mathematics and the biology slope coefficients is 0.17 (1-0.83) and is statistically significant. The interpretation of this difference is that students' relative proficiency is more easily signalled in mathematics than in biology.[47] The high school grading policy slope coefficients are much smaller than 1. The admission standards, combined with bell-shaped university grading, can explain the difference in university and high school slope coefficients.[48] The difference between the university mathematics and the high school intercepts is about 77 points. This difference captures the greater difficulty of university

[45] Loury and Garman (1995) look at weekly earnings while Jones and Jackson (1990) look at annual earnings. Krashinsky (2006) estimates the impact of Grade 13 on earnings assuming that a one-point increase in GPA translates into a 10% earnings increase.

[46] In order to get this estimate I also assumed that an increase in 2.2 points would lead to a 0.22 increase in GPA since letter grades contain 10 points.

[47] Many factors could explain this difference. For example, the test formats are different: biology test questions are all multiple-choice questions while mathematics uses a mixture of question types. Because of the nature of the multiple-choice questions, luck might play a bigger role, relative to ability, in biology than in mathematics for lower ability students.

[48] Students admitted to the university have high school averages above 80%. At the university level, we usually observe grades varying between 30 and 100%. So, for accepted students, the span of grades is increased between high school and university while the span of ability is fixed. As a consequence, the payoff of an extra unit of ability has to be more important at the university level to cover the new span of grades.

courses and the more intense competition in university classrooms.

The average levels of initial academic ability seem to differ across groups. This finding, combined with the results suggesting an ability measurement discrepancy across subjects, favors the use of the grading policy model over difference-in-differences estimation. I test the significance of this difference by re-parametrizing the model.[49] The estimated difference $\hat{\Delta}_\eta$ is -1.77, which is just the difference between estimated levels of initial academic ability ($\bar{\eta}^{G13}$ and $\bar{\eta}^{G12}$) obtained in Table 7. Given the standard error of $\hat{\Delta}_\eta$ is 0.08, I can reject the hypothesis of equal average ability. G12 students look brighter than G13 students, which is consistent with the selection story in which brighter G13 students escaped from the double cohort (and from the sample).

Results from high school grading policies do not reveal any clear pattern in the way teachers graded students in high school. Even though the intercept coefficient of the high school grading policy for G13 students is more important than for G12 students, the opposite is true for the slope coefficient. Using LR tests, I successively test for the equality of slope coefficients and the equality of the intercept coefficients. Table 8 presents the results.

Table 8: Testing for the Equality of High School Grading Policies

	d	df	p-value	Δd	df	p-value
Baseline model*	0.18	1	0.67	-	-	-
Imposing:						
Same slope coefficients	0.92	2	0.63	0.74	1	0.39
Same intercept coefficients	1.48	2	0.48	1.30	1	0.25
Same slope & intercept coefficients	177.5	3	0.00	177.3	2	0.00

* The baseline model is the model used to present the results in Table 7.

I test the restrictions of equal slope and intercept coefficients by comparing the fit of the restricted models to the fit of the model used to measure the value-added of G13 (labeled as the baseline model). The first step in doing so is to test whether the model presented in Table 7 fits the data well. If the model is valid, then N times the fit function evaluated at the estimated coefficient values ($d = NF(\hat{\theta})$) is asymptotically χ^2 distributed: the closer d is to 0, the better the fit. The model used to measure the value-added of G13 fits the data well, given to the low value of d. In the case of this model, the test simply looks at whether the ratio of covariances in equation (15)

[49] Instead of fixing the intercept term of the G12 mathematics grading policy to zero, as was done above, I fix the average initial level of general academic ability for the G12 to zero ($E_{G12}[\eta_i] = 0$). This way, the estimate of $E_{G13}[\eta_i]$ will give us the difference in ability, with correct standard errors.

is the same for both groups since there is only one overidentifying restriction. The second step is to compare the fit of the restricted models I want to test with the fit of the baseline model. If the restrictions imposed on the model are valid, then the difference in fit (between the restricted and unrestricted models) measured by Δd is also asymptotically χ^2 distributed.[50] The p-values of these tests are presented in the last column of Table 8. I cannot reject the hypothesis that both high school grading policies have the same slope or intercept coefficient, but I do reject the hypothesis that the grading policies are the same (equal slope and intercept coefficients). The results from the tests might seem surprising but we have to remember that the variation in high school marks is small and that no mark is close to zero. As a consequence, it is almost impossible to disentangle a small shift in intercept from a small shift in slope coefficients.

8.1 Summary of Results

The results all suggest that the value-added to Grade 13 is modest for students who will attend university. Estimates of the value-added are similar whether I use means comparison (0.45), OLS estimation (1.68), difference-in-differences (1.92), or the grading policy model (2.21) as way of capturing the value-added of Grade 13.

That said, the factor model proves to be useful in capturing effects which the other methods presented in the paper do not account for. The results from the factor model show that difference-in-differences estimation would lead to biased estimates of the value-added of Grade 13 if biology and mathematics do not measure ability in the same way. In the present case, the factor model estimate is 15% above the difference-in-differences estimate and 32% above the OLS estimate. The factor model estimate is above the OLS estimate because of the correlation between the amount of schooling and the average level of ability of students. It is also close to five times the means difference estimate, which shows the importance controlling for heterogeneity in average ability level across the two groups.

[50] The number of degrees of freedom of Δd is given by the difference in degrees of freedom of the compared models. See Chamberlain (1984) for details.

8.2 Heterogeneity

When compared to previous studies, results presented here suggest the presence of substantial heterogeneity in the return to an extra year of high school. There are at least two ways to check for heterogeneity in the value-added of Grade 13 within the present sample. I first separated the sample in two groups based on their academic ability. I formed a higher-ability group and a lower-ability group using the median university biology grade as a cutoff point. Estimating the value-added separately for each group, I find that the value-added for lower-ability students is 1.4 points greater than for the higher-ability students. Although modest, the difference has the expected sign – lower-ability students gain more from an extra year of high school. Alternatively, I can introduce an extra parameter in the grading policy model since there is one degree of freedom in the baseline model. For example, it is easy to rewrite equations (7) and (8) to allow for the value-added to be a linear function of students' ability:

$$\eta_i^{U,B} = \eta_i(1+\phi(\tau^{H,B}+\tau^{U,B}))+\tau^{H,B}+\tau^{U,B} \tag{20}$$

$$\eta_i^{U,M,C} = \eta_i(1+\phi(\tau^{H,M,C}+\tau^{U,M}))+\tau^{H,M,C}+\tau^{U,M} \tag{21}$$

where ϕ is the heterogeneity coefficient. If the value-added is decreasing with ability we would expect ϕ to be negative. Note that equations (20) and (21) are identical to (7) and (8) if there is no heterogeneity (i.e. $\phi = 0$).

I re-estimated the model allowing for heterogeneity in the value-added (as specified by equations (20) and (21)) and found that $\hat{\phi} = -0.03$. The estimate is not statistically significant, as we could guess from the estimate of d for the baseline model in Table 8. Overall, it is still surprising to find some evidence of heterogeneity in the value-added of an extra year of high school for such a homogeneous group of individuals.

9 Robustness

This section considers the robustness of the main results.

The results presented in Table 7 assumed that female and male students get the same benefit out of Grade 13. However, this might not be the case. In order to investigate the possibility of heterogeneity across gender in the value-added of Grade 13, I estimated the grading policy model separately for females and males. Table 9 presents the results by gender.

Table 9: Parameter Estimates by Gender

	Females		Males	
	G13	G12	G13	G12
υ (intercept)				
MATH	2.43	(-)	1.97	(-)
	(0.08)	(-)	(0.20)	(-)
BIO	15.85	15.85	19.36	19.36
	(3.21)	(3.21)	(5.15)	(5.15)
HS	78.38	77.02	76.98	75.18
	(1.31)	(1.25)	(2.45)	(1.65)
λ (slope)				
MATH	1.00	1.00	1.00	1.00
	(-)	(-)	(-)	(-)
BIO	0.86	0.86	0.79	0.79
	(0.05)	(0.05)	(0.07)	(0.07)
HS	0.19	0.20	0.19	0.22
	(0.02)	(0.02)	(0.03)	(0.02)
$\bar{\eta}$ (ability)	66.97	68.42	71.04	73.31
	(0.08)	(0.01)	(0.20)	(0.01)
N	291	323	145	179

Table 9 does not suggest that the results presented in Table **??** are driven by a specific gender. The parameter estimates are similar across genders. Both estimation results suggest that the value-added of Grade 13 is modest and that G12 and G13 students are different in terms of initial levels of ability.

I replicated the experiment using chemistry instead of biology. Chemistry is another course that life science students must take which was not affected by the reform and for which a student's performance should not be influenced by her mathematics knowledge. The results are similar to the ones presented here (the estimated value-added of G13 is 1.7 points). I also replicated the estimates using chemistry instead of mathematics. In this case, any evidence of value-added of Grade 13 would be problematic. The estimated value of Grade 13 ($\hat{\Delta}_V$) in this case is very small (0.25). This evidence supports the hypothesis that biology and chemistry were not affected by the reform.

Because covariances are sensitive to outliers, I did not include 12 students with grades below 30% in either biology or mathematics and assumed that these students dropped out. Including these students does not change the results ($\hat{\Delta}_V = 2.17$ as compared to 2.21). Also, students only get a grade if they complete the course they are enrolled in. If a disproportionate fraction of G12 students drop out of mathematics, then the G13 value-added estimator would be biased. Interestingly, there are no students who officially dropped out of mathematics but who completed biology. This could be due to the fact that these courses are compulsory for admission into life sciences specialization fields. When we look at the unconditional drop-out rates in these two courses, we realize that they are similar, and for both courses relatively low (5% for mathematics and 2% for biology).

G12 students could take fewer courses if they feel less well prepared than G13 students to face university challenges. This is not the case. G12 students take an average of 5.8 courses over the first year while G13 take 5.7 from the Faculty of Arts and Science. The difference is very small.

Students also select the program they want to attend. G12 students, perhaps knowing that their preparation in mathematics is not as good as G13, might have avoided applying to programs involving mathematics. But students do not differ significantly in terms of the program they chose (within the Faculty of Arts and Science). In fact, there is a slightly larger proportion of G13 students who chose a humanities over a life science program than G12, which again supports the hypothesis that G12 did not try to compensate for their lack of mathematics preparation.[51]

If Grade 13 gives students general human capital that affects all subjects similarly, then the estimation methods presented in this paper would fail to capture the full extent of the benefit of this extra year. This would be true if, for some reason, the reform affected a student's university biology performance as well as mathematics performance. It is true that schooling might bring more to students than just specific knowledge – for instance, maturity gained while being in school. If so, G13 students would be expected to do better than G12 students in every course. Table 10 suggests otherwise:[52] G12 students do not do significantly worse than G13. They actually do better in a majority of courses (except mathematics). I cannot totally rule out the possibility of such an effect since the higher average ability level of G12 could compensate for the lack of maturity, for example: the maturity effect could be confounded with academic ability as defined in this paper.

A sign of such a missing variable could be the presence of group-heteroskedasticity. For ex-

[51] Humanities represent 38% of G12 students and 41% of G13 students applications. Life-Science represents 39% of G12 students and 34% G13 students applications.

[52] The subjects analyzed in Table 10 are anthropology, biology, chemistry, economics, history, mathematics (for business and life sciences), philosophy, psychology, and sociology.

Table 10: Students Average Marks in 2003

	# Obs.	Marks		$H_0: G13 = G12$
		G13	G12	p-values
ANT100	501	**66.8**	65.7	0.34
BIO150	1161	73.2	**74.4**	0.08
CHM138	1016	76.2	**76.8**	0.47
ECO100	629	68.1	**68.2**	0.91
HIS109	293	67.2	**69.0**	0.27
MAT133	281	**69.8**	67.3	0.14
MAT135	1092	**68.5**	68.5	0.97
PHL100	448	71.4	**72.0**	0.49
PSY100	883	70.0	**70.3**	0.73
SOC101	791	**65.5**	64.7	0.30

ample, if the G13 students are mature while only some G12 students are, then maturity should not play a role in the within-group grades variation for G13 grades but should play a role for G12 students. We would expect the two groups to have different variability in the error terms (since the variability in maturity would be included in the variance of the error term). The null hypothesis of homoskedasticity cannot be rejected using a similar test used for the equality of high school grading policies.[53] Without ruling out the possibility of such a general effect on students, it is hard to find support for such an effect with the data I analyzed. If we think that high school teachers are grading students similarly across groups (and remember that I cannot reject the hypothesis of equal slope or intercept high-school coefficients), it would be even harder to support such a possibility.

Effort may be a factor influencing students' performance.[54] If the amount of effort is the same in both groups (G12 and G13) or if it is constant across courses for the same group then effort should not affect the validity of my results. In the first case, it would not affect the groups' relative performance, while in the second case, the difference in effort level would be captured by the difference in the average ability measure. But students can use effort to compensate for their lack of preparation in mathematics. G12 students might put more effort into studying mathematics than G13 students. If there is an important substitution effect between study time for mathematics and study time for biology, then the effect of Grade 13 would be diluted by the extra effort exerted by G12 students in mathematics, and the estimate of the value-added of G13 would then be downward-

[53] $\Delta d = 1.93$. The p-value is 0.38.

[54] See Stinebrickner and Stinebrickner (2005) for an IV estimation of the impact of time allocation on students' college performance.

biased. The substitution effect would influence both the mathematics and biology grades. This means that the difference in ability would also be downward-biased (since the performance of G12 students in biology would be negatively affected).

The absence of information about students' study habits makes it impossible to formally test for the presence of effort substitution. But Table 10 does not suggest the presence of such behavior from G12 students. If G12 students substituted effort from biology or chemistry to mathematics then we would expect to see the difference in performance between the two groups being more important for courses in which students are not expected to take any advanced mathematics. Humanities subjects should favor more G12 students than biology, chemistry, or economics. This is not the case. Anthropology, history, philosophy, and sociology, as a whole, do not favor G12 more than biology, chemistry and economics. Overall, there is no strong evidence that the factor model measure of the value-added of Grade 13 is downward-biased.

10 Conclusion

The 1999 Ontario Secondary School reform provides a valuable opportunity to estimate the return to a year of secondary schooling for students who will pursue post-secondary education, a population that represents a large portion of students in most developed countries. The results obtained in this study suggest that the return to schooling for these students is modest. In particular, I find that students coming out of Grade 13 only have a 2.2 point advantage (on a 100 point scale) over students from Grade 12, once I control for ability differences. The estimated return to schooling is around 2 percent.

These results contrast with findings from previous studies examining the returns to schooling. The magnitude of the return to Grade 13 is modest compared to previous LATE estimates, especially compared with estimates found in studies looking at the impact of schooling on potential high-school dropouts. Whether I use means comparison, OLS estimation, difference-in-differences, or the factor model presented above, the returns to schooling never reach returns comparable to the ones found in the previous IV literature. But even if the results suggest returns far below the usual estimated return to schooling, these results are by no means in conflict with that earlier literature. Indeed, they support Lang and Card's LATE interpretation of the IV estimates of the return to schooling, representing solid evidence of heterogeneity in the return to schooling across ability levels. My findings contrast with Krashinsky (2006), who looks at the impact of the same reform on a population of students with lower high school averages. He finds that the return to Grade 13 for these students is between 6 and 12 percent, which contrasts with this paper's 2.2 percent further supporting the heterogeneity in the return to secondary education.[55] Whether the modest value-added of G13 found in this paper will affect students in the longer-run is an issue that warrants further investigation.

References

[1] AAVIK, A., K. G. SALVANES, AND K. VAAGE (2003): "Measuring Heterogeneity in the Returns to Education in Norway Using Educational Reforms," IZA Discussion Paper No. 815.

[2] ABOWD, J. M., AND D. CARD (1987): "Intertemporal Labor Supply and Long Term Employment Contracts," *American Economic Review*, 77, 50-68.

[3] ABOWD, J. M., AND D. CARD (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57, 411-445.

[4] AIGNER, J. A. *et al.* (1984): "Latent Variable Models in Econometrics," in Z. Griliches and M. Intrilligator Eds., *Handbook of Econometrics.*

[5] ALTONJI, J. G., AND L. M. SEGAL (1996): "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, 14, 353-366.

[6] ALTONJI, J. G., A. P. MARTINS AND A. SIOW. (2002): "Dynamics Factor Models of Consumption, Hours and Income," *Research in Economics*, 56, 3-59.

[7] ANGRIST, J. D. AND A. B. KRUEGER (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 106 (4), 979-1014.

[8] ANGRIST, J. D. AND A. B. KRUEGER (1999): "Empirical Strategies in Labor Economics," in O. Ashenfelter and D. Card eds., *Handbook of Labor Economics*, Vol. 3A, 1277-1366.

[55] Also, while the high-ability students do not seem to have been severely affected by the OSS reform, King *et al.* (2002, 2004, 2005) report that lower ability students were adversely affected by the curriculum compression. King *et al.* (2004) note when talking about workplace-bound students' credit accumulation toward high school graduation: "These data suggest there are serious problems with the progress of students taking Applied courses."

[9] ANGRIST, J. D. AND V. LAVY (1999): "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 114 (2), 533-575.

[10] BAKER, M. (1997): "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings," *Journal of Labor Economics*, 15, 338-375.

[11] BOLLEN, K. A. (1989). Structural Equations with Latent Variables. Wiley, New York.

[12] CANADIAN EDUCATION STATISTICS COUNCIL (2003), "Education Indicators in Canada: Report of the Pan-Canadian Educational Indicators Program," Catalogue no. 81-582-XPE, Statistics Canada.

[13] CARD, D. (1995): "Earnings, Schooling and Ability Revisited," in Solomon W. Polachek, ed., *Research in Labor Economics*, Vol. 14, 23-48.

[14] CARD D. (1999): "Causal Effect of Education on Earnings," in O. Ashenfelter and D. Card eds., *Handbook of Labor Economics*, Vol. 3A, 1801-1863.

[15] CARD, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69, 1127-1160.

[16] CHAMBERLAIN, G. (1984): "Panel Data," in Z. Griliches and M. Intrilligator Eds., *Handbook of Econometrics*.

[17] CUNHA, F., J. J. HECKMAN, L. LOCHNER, AND D. V. MASTEROV (2005): "Interpreting the Evidence on Life Cycle Skill Formation," NBER working paper no. 11331.

[18] HARMON, C., AND I. WALKER (1995): "Estimates of the Economic Return to Schooling for the United Kingdom," *American Economic Review*, 49, 1377-1398.

[19] HAYDUK, L. A. (1989): "Structural Equation Modeling with LISREL: Essentials and Advances," John Hopkins University Press, Baltimore, 405 p.

[20] HOXBY, C. M. (2000): "The Effects of Class Size on Student Achievement: New Evidence from Natural Population Variation," *Quarterly Journal of Economics*, 115 (4), 1239-1286.

[21] IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-475.

[22] JONES E. B., AND J. D. JACKSON (1990): "College Grades and Labor Market Rewards," *Journal of Human Resources*, 25, 253-266.

[23] KING, A. J. C. *et al.* (2002): "Double Cohort Study: Phase 2 Report," Social Program Evaluation Group, Queen's University.

[24] KING, A. J. C. *et al.* (2004): "Double Cohort Study: Phase 3 Report," Social Program Evaluation Group, Queen's University.

[25] KING, A. J. C. *et al.* (2005): "Double Cohort Study: Phase 4 Report," Social Program Evaluation Group, Queen's University.

[26] KRASHINSKY, C. (2006): "How Would One Extra Year of High School Affect Academic Performance in University? Evidence from a Unique Policy Change," Working Paper.

[27] KRUEGER A. B. (1999): "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics*, 114 (2), 497-532.

[28] LANG, K. (1993): "Ability Bias, Discount Rate Bias and the Return to Education," Mimeo. (Department of Economics, Boston University).

[29] LITTLE, T. D. (1997): "Mean and Covariance Structures (MACS) Analysis of Cross-Cultural Data: Practical and Theoretical Issues," *Multivariate Behavioral Research*, 32 (1), 53-76.

[30] LOURY, L. D., AND D. GARMAN (1995): "College Selectivity and Earnings," *Journal of Labor Economics*, 13, 289-308.

[31] MEGHIR, C., AND M. PALME (2005): "Educational Reform, Ability, and Family Background," *American Economic Review*, 95, 414-424.

[32] MINCER, J. (1974): Schooling, Experience, and Earnings, New York: NBER Press.

[33] ONTARIO MINISTRY OF EDUCATION, (1984): "Ontario Schools: Intermediate and Senior Divisions, Program and Diploma Requirements."

[34] ONTARIO MINISTRY OF EDUCATION, (1985), "Mathematics, Intermediate and Senior Divisions, Curriculum Guideline. Part 3: Mathematics."

[35] ONTARIO MINISTRY OF EDUCATION, (1987), "Science, Intermediate and Senior Divisions, Curriculum Guideline. Part 12: Biology."

[36] ONTARIO MINISTRY OF EDUCATION AND TRAINING, (1999): "Ontario Secondary Schools, Grades 9 to 12: Program and Diploma Requirements."

[37] ONTARIO MINISTRY OF EDUCATION AND TRAINING, (2000a): "The Ontario Curriculum Grades 11 and 12: Mathematics."

[38] ONTARIO MINISTRY OF EDUCATION AND TRAINING, (2000b): "The Ontario Curriculum Grades 11 and 12: Science."

[39] ONTARIO UNIVERSITIES' APPLICATION CENTRE Website, (2006), "Undergraduate Application Statistics," accessed June 1^{st} 2006. http://www.ouac.on.ca/news/uapp/underappstats.html

[40] OREOPOULOS, P. (2006): "Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter," *American Economic Review*, 96, 152-175.

[41] PLOYHART, R. E., AND F. L. OSWALD (2004): "Applications of mean and covariance structure analysis: Integrating correlational and experimental approaches," *Organizational Research Methods*, 7 (1), 27-65.

[42] STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557-586.

[43] STINEBRICKNER T. R. AND R. STINEBRICKNER (2005): "The Causal Effect of Studying on Academic Performance," working paper.

[44] WANSBEEK, T. J., AND E. MEIJER, (2000): Measurement errors and latent variables in econometrics. Advanced Textbooks in Economics, 37. North-Holland, Amsterdam.

A Appendix

A.1 Coefficients Identification

Here I simply show one of the different possible strategies. I start by expanding the basic hypotheses of general structural equation models ($\Sigma^{(g)} = \Sigma(\Theta)^{(g)}$, and $\mu^{(g)} = \mu(\Theta)^{(g)}$)

$$
\begin{aligned}
var(H_i^C) &= \left[\lambda^{HC}\right]^2 \sigma^2_{\eta^C} + \sigma^2_{\varepsilon^{H,C}} && (22)\\
var(B_i^C) &= \left[\lambda^{B}\right]^2 \sigma^2_{\eta^C} + \sigma^2_{\varepsilon^{B,C}} && (23)\\
var(M_i^C) &= \sigma^2_{\eta^C} + \sigma^2_{\varepsilon^{M,C}} && (24)\\
cov(H_i^C, B_i^C) &= \lambda^{HC}\lambda^{B}\sigma^2_{\eta^C} && (25)\\
cov(H_i^C, M_i^C) &= \lambda^{HC}\sigma^2_{\eta^C} && (26)\\
cov(B_i^C, M_i^C) &= \lambda^{B}\sigma^2_{\eta^C} && (27)\\
E(H_i^C) &= \upsilon^{HC} + \lambda^{HC} E_C\left[\eta_i\right] && (28)\\
E(B_i^C) &= \upsilon^{B} + \lambda^{B} E_C\left[\eta_i\right] && (29)\\
E(M_i^{G13}) &= \Delta_V + E_{G13}\left[\eta_i\right] && (30)\\
E(M_i^{G12}) &= E_{G12}\left[\eta_i\right] && (31)
\end{aligned}
$$

The identification of the average academic ability for the G12 students ($E_{G12}\left[\eta_i\right]$) is trivial from the normalization I made. Next, we can isolate λ^B using (25) and (26)

$$\lambda^B = \frac{cov(H_i^C, B_i^C)}{cov(H_i^C, M_i^C)} \tag{32}$$

Dividing (25) by (27) we get

$$\lambda^{HC} = \frac{cov(H_i^C, B_i^C)}{cov(B_i^C, M_i^C)} \tag{33}$$

Equations (32), (31), and (29) give us an expression for the constant term of the biology grading policy

$$
\begin{aligned}
\upsilon^B &= E(B_i^{G12}) - \lambda^B E_{G12}\left[\eta_i\right]\\
&= E(B_i^{G12}) - E(M_i^{G12})\frac{cov(H_i^C, B_i^C)}{cov(H_i^C, M_i^C)} && (34)
\end{aligned}
$$

Plugging (34) and (32) in (29) will give us a measure of the average academic ability of G13 students

$$\begin{aligned} E_{G13}\left[\eta_i\right] &= \frac{E(B_i^{G13}) - \upsilon^B}{\lambda^B} \\ &= \frac{cov(H_i^C, M_i^C)}{cov(H_i^C, B_i^C)}\left[E(B_i^{G13}) - E(B_i^{G12})\right] + E(M_i^{G12}) \end{aligned} \tag{35}$$

Having isolated the average academic ability of G13 students, I am able to identify the value-added

$$\begin{aligned} \Delta_V &= E(M_i^{G13}) - E_{G13}\left[\eta_i\right] \\ &= E(M_i^{G13}) - E(M_i^{G12}) - \frac{cov(H_i^C, M_i^C)}{cov(H_i^C, B_i^C)}\left[E(B_i^{G13}) - E(B_i^{G12})\right] \end{aligned} \tag{36}$$

The constant term of G13 students high school grading policy can be found using (35), (33), and (28)

$$\begin{aligned} \upsilon^{HG13} &= E(H_i^{G13}) - \lambda^{HG13} E_{G13}\left[\eta_i\right] \\ &= E(H_i^{G13}) \\ &\quad - \frac{cov(H_i^{G13}, B_i^{G13})}{cov(B_i^{G13}, M_i^{G13})}\left\{E(M_i^{G12}) + \frac{cov(H_i^C, M_i^C)}{cov(H_i^C, B_i^C)}\left[E(B_i^{G13}) - E(B_i^{G12})\right]\right\} \end{aligned}$$

while the same constant for G12 students looks like

$$\begin{aligned} \upsilon^{HG12} &= E(H_i^{G12}) - \lambda^{HG12} E_{G12}\left[\eta_i\right] \\ &= E(H_i^{G12}) - \frac{cov(H_i^{G12}, B_i^{G12})}{cov(B_i^{G12}, M_i^{G12})} E(M_i^{G12}) \end{aligned}$$

A.2 OLS Estimation Bias

This section shows conditions under which the OLS regression estimation of mathematics grades on biology grades gives biased estimator of the G13 value-added . I first transform the data in deviations from mean and work with matrix notation for clarity reasons

$$\tilde{M}_i = \Delta_V \tilde{I}_i + \lambda^B \tilde{B}_i + u_i$$

or

$$\tilde{M} = \tilde{X}\beta + u$$

where $\tilde{M}$ is a vector of demeaned mathematics grades and $\tilde{X}$ is the matrix of the now demeaned variables. The OLS estimator of β, $\hat{\beta}$, is

$$(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{M}$$

Hence $plim\ \hat{\beta}$ is

$$plim\ \hat{\beta} = \beta + plim(\frac{\tilde{X}'\tilde{X}}{N})^{-1} plim\left(\frac{\tilde{X}'u}{N}\right)$$

The dummy variable indicating if the student is a member of the G13 group is not correlated with the error, the first term of $plim\left(\frac{\tilde{X}'u}{N}\right)$ will be 0. But since $\tilde{B}_i$ is correlated with the error term (because of measurement error) the second term of the vector will be non-zero. This gives

$$plim\left(\frac{\tilde{X}'u}{N}\right) = \begin{bmatrix} 0 \\ -\frac{var(\varepsilon_i^B)}{\lambda^B} \end{bmatrix}$$

Unless we are facing a $plim(\frac{\tilde{X}'\tilde{X}}{N})^{-1}$ with 0 off-diagonal elements (e.g. the regressors are uncorrelated) we should get a biased estimator of the treatment effect.

$$plim(\frac{\tilde{X}'\tilde{X}}{N})^{-1} = plim\begin{bmatrix} \frac{N_{G12}N_{G13}}{N^2} & \sum_{i\in G13}^{N_{G13}} \frac{\tilde{B}_i}{N} \\ \sum_{i\in G13}^{N_{G13}} \frac{\tilde{B}_i}{N} & \sum^{N} \frac{\tilde{B}_i^2}{N} \end{bmatrix}^{-1}$$

Let

$$\begin{aligned} plim(\frac{\tilde{X}'\tilde{X}}{N})^{-1} &\equiv plim\,(A)^{-1} \\ &= plim\frac{1}{\det A} plim\begin{bmatrix} \sum^{N} \frac{\tilde{B}_i^2}{N} & -\sum_{i\in G13}^{N_{G13}} \frac{\tilde{B}_i}{N} \\ -\sum_{i\in G13}^{N_{G13}} \frac{\tilde{B}_i}{N} & \frac{N_{G12}N_{G13}}{N^2} \end{bmatrix} \\ &\equiv plim\frac{1}{\det A} plim\begin{bmatrix} B & -C \\ -C & D \end{bmatrix} \end{aligned}$$

Hence

$$plim\ \hat{\beta} - \beta = plim \frac{1}{\det A} plim \begin{bmatrix} B & -C \\ -C & D \end{bmatrix} \begin{bmatrix} 0 \\ -\frac{var(\varepsilon_i^B)}{\lambda_A} \end{bmatrix}$$

As long as $C \neq 0$, the OLS estimator of the G13 value-added will be biased. C could be equal to 0 if the average grades for both groups are equal which could happen if both groups had the same level of academic ability. Then the dummy variable would be uncorrelated with the student grades in biology. The sign of the bias will depend on whether the average for G13 students is higher or lower than G12 students in biology. If the average is higher for the latter group, we should expect a downward bias.