C.I.R.P.É.E

Centre Interuniversitaire sur le Risque, les Politiques Économiques et l'Emploi

Cahier de recherche/Working Paper **08-30**

Family Background, Family Income, Cognitive Tests Scores, Behavioural Scales and their Relationship with Post-secondary Education Participation: Evidence from the NLSCY

Pierre Lefebvre

Philip Merrigan

Septembre/September 2008

Lefebvre/Merrigan : Economics, School of Management, Université du Québec à Montréal, CP 8888, Succ. Centre-ville, Montréal, QC, Canada H3C 3P8 ; and CIRPÉE
lefebvre.pierre@uqam.ca
merrigan.philip@uqam.ca

This analysis is based on Statistics Canada's National Longitudinal Survey of Children and Youth (NLSCY) restricted-access Micro Data Files, which contain anonymized data collected in the NLSCY and are available at the Québec Inter-university Centre for Social Statistics (QICSS), one of the Canadian Research Data Centers network. All computations one these micro-data were prepared by the authors who assume the responsibility for the use and interpretation of these data. This research was funded by Human Resources and Social Development Canada. We thank David Groux for his excellent research assistance and participants to Statistics Canada Socio-economic Conference 2008, May 5-6, 2008, Ottawa Congress Centre; MESA Session/Canadian Economic Association Meeting 2008, June 7-8, UBC, Vancouver.

Abstract:
This paper exploits the panel feature of the Canadian National Longitudinal Survey of Children and Youth (NLSCY) and the large diversity of measures collected on the children and their families over 6 cycles (1994-1995 to 2004-2005) to explain high school graduation and postsecondary education (PSE) choices of Canadian youth aged 18 to 21 observed in the most recent wave of the survey. In estimating how family background, family income, cognitive abilities, non-cognitive abilities and behavioural scores influence schooling choices they can used as markers for identifying children at risk of not pursuing PSE. We focus on the impact of measures that are specific to the NLSCY which contains a host of scores on several dimensions such as the cognitive achievement of children (reading and math test scores); behavioural scores that measure the levels of hyperactivity, aggression, and pro-sociality; scores that measure self-esteem and self-control (non-cognitive abilities); and, finally scores that measure the quality of parenting, family dysfunction, of neighbourhoods and schools quality. The math and reading scores are particularly interesting because they are computed from objective tests and are not based on any type of recall, as compared, for example, with the Youth in Transition Survey (YITS) data set.

Despite the fact that income, as measured as the mean income ($2002) of the family during cycles 1 to 4, does not seem to be a key player for PSE attendance or high school graduation, the sign of its effect is generally positive and non-linear, increases for children in very low income will have a larger effect that those with higher levels. More importantly, several variables that are characteristics of low-income families play a key role for schooling attainment. For example, being from a single-parent/guardian home with a poorly educated PMK and with less than (perceived) excellent/very good health or with high levels of hyperactivity for males or high levels of aggression for young teenage females will almost negate any chance of attaining the level of PSE.

Keywords: High school graduation, postsecondary education, schooling transition, gender, youth, longitudinal data

JEL Classification: I21, J13, J16, J24

1. Introduction

There is considerable evidence that, on average, youth from lower income family participate less in postsecondary education (PSE) and have lower educational attainment than their better-off peers. Recent studies in child development, using panel non-experimental data, insist on the dynamic nature or cumulative process at different period (stages) of the life cycle producing achievements (Todd and Wolpin, 2007; Cunha, 2007; Cunha and Heckman, 2007; Cunha et al., 2006; Heckman, 2007; Ermisch and Francesconi, 2001). These studies suggest that the shaping of skills and educational attainment of children is intimately related to the family environment (investments, resources, transmitted skills, values, motivation, etc.). Therefore, a particularly rich set of data is necessary to address the topic of PSE. The purpose of this research is to estimate the relationship between family background, family income, cognitive and non-cognitive test scores, behavioural scales and educational attainment for young men and women in Canada. The specific objectives are:

1) Exploit the panel feature of the Canadian National Longitudinal Survey of Children and Youth (NLSCY) and the large diversity of measures collected on the children and their families (6 cycles of data from 1994-1995 to 2004-2005) to explain schooling transitions and choices of Canadian youth aged 18 to 21 who were last observed in the most recent wave (Cycle 6, 2004-2005) of the survey.
2) Estimate how family background, family income, cognitive abilities, non-cognitive abilities as defined in Cunha and Heckman (2006), behavioural scores cause changes in schooling choices and can be used as markers for identifying children at risk of not pursuing PSE; we will be focusing on the impact of measures that are specific to the NLSCY which contains a host of scores on several dimensions such as the cognitive achievement of children (reading and math test scores); behavioural scores that measure the levels of hyperactivity, aggression, and pro-sociality; scores that measure self-esteem and self-control (non-cognitive abilities); and, finally scores that measure the quality of parenting, family dysfunction, and of neighbourhoods. The math and reading scores are particularly interesting because they are computed from objective tests and are not based on any type of recall, as compared, for example, with the YITS data set.
3) Measure the total impact (including the impact on the probability of HS graduation) of these backgrounds family/parental/youth characteristics on the probability of choosing a particular type of PSE (College or University) education according to postsecondary institutional settings (Québec, and the other provinces), dropping-out or not pursuing a PSE degree after HS graduation.
4) Derive policy implications from the results for both federal and provincial levels of government.

We shall derive the estimated impact of these different factors using regression analysis. We will start out by estimating a model explaining the probability of graduating from high school for the cohort of 18 to 21 year-old youth in cycle 6. All regressions are done with a sample of young men and young women separately. We shall explain later why in the paper. We then estimate, given that they have graduated from high school, the probability youth attain the PSE level and with a multinomial Logit the probability they are HS graduates who do not pursue PSE, college or university attendees. We end our regression analysis with a Sequential Logit. This procedure estimates simultaneously the probability of graduation from HS, as well as the probability of stopping after HS. We then compute the effect of key variables on the overall probability of PSE

attendance which is the probability of graduating from HS times the probability of PSE attendance. Finally, we assess the bias from estimating a Sequential Logit assuming that the unobservable variables that determine graduation from HS and those that determine PSE attendance are correlated. We find that the bias is small and that the Sequential Logit presents a lower bound on the effects of these key variables. However, we do find evidence that there is no correlation between unobservable variables that determine both HS graduation and PSE attendance. Given that the models with correlation take considerable time to estimate, we do policy analysis with the results from the Sequential Logit which are precise enough, we believe, to warrant a serious policy discussion.

The structure of the paper is the following. Section 2 starts out with a review of the recent empirical evidence on PSE attendance. Section 3 briefly reviews the modelling strategies employed by researchers to analyse the links between education attainment and family background. Section 4 describes the information collected in the NLSCY, the data set available for cycle 6 respondents, and the variables constructed and used specifically for estimation. The estimations results are presented and discussed in section 5. Section 6 discusses the self-selection problem and bias that may occur in our estimations and presents the Sequential Logit results. Section 7 derives Policy implications from the results. Section 8 summarizes and concludes the paper.

2. Review of recent research

In Canada, a few studies from diverse of data sets (e.g. Survey on Consumer Finances/SCF, Survey of Labour and Income Dynamic/SLID, Youth in Transition Survey/YITS) have analyzed the link between participation in post-secondary education (PSE) and parental income (Knighton and Mirza, 2002; Corak et al., 2003; Cristofides et al., 2001; Frenette, 2007; Drolet, 2005; Rahman et al., 2005). They show that participation rates are higher among youth from high-income families with more educated parents. They also provide evidence that the effects of family income have not varied in the late 90's and early twenty first century which shows that the increase in university fees across Canada over the nineties have not decreased PSE attendance.

On the other hand, transitory family income shocks could be a crucial factor according to the results of Coelli (2005). Tuition fees (given the across-province and time variation of fees observed in Canada) or financial constraints do not seem to be an important determinant of participation (Christofides et al., 2001; Frenette, 2007; Rivard and Raymond, 2004). Family background characteristics (such as parents' education and family structure) play a larger role as well as grade point average in high school for postsecondary participation (preceding cited studies, Finnie et al. 2004).

Statistics Canada's longitudinal Youth in Transition Survey (YTIS) collects unique information on the educational and labour market pathways of a sample of young Canadians surveyed at the ages of 18 to 20 in 1999 (YTIS-B Cohort) and at age 15 in 2000 (YTIS-A Cohort).[1] They have been interviewed three additional times since then (2002, 2004, and 2006).

The first panel of Table 1 summarizes descriptive statistics presented by Shaienks, Eisl-Culkin and Bussière (2006), from cycles 1 to 3 for the B-Cohort that we complement with cycle 4 statistics. It highlights an important fact about participation in secondary education: although the usual age of graduation from high school is 18 in most provinces (17 in Québec), approximately 76% of youth received their diploma in December 1999 at the ages of 18 to 20 years; another 13% are high-school continuers and manage to graduate four years later (including high school dropouts). Two years later, 2% more have graduated, reducing the percentage to 8%. Again, a large gender gap exists at ages 18 to 20, which decreases over time for participation in secondary education but manifests itself for PSE participation (as shown in panel 2 of Table 1).

Shaienks and Gluszynski (2007) use the four cycles to describe the schooling trajectories of B-Cohort respondents until they are 24 to 26 years of age. The second panel of Table 1 shows that the proportion of youth who participate in PSE increases substantially with age and appears to level off at 24-26. Concurrently, dropping out of postsecondary education increases with age (15% by December 2005). University participation and the proportion of PSE graduates who continue are still increasing. A significant gap between young females and males can be observed for all education statuses related to PSE.

[1] A series of research projects exploiting these data sets are on-going financed by the Canadian Millennium Scholarship Foundation through the MESA project, an undertaking based at the School of Policy Studies at Queen's University. To date very few of the working paper are available.

Table 1: Education participation rate of youth [women] (men) by level, YTIS by December 1999, 2001, 2003 and 2005

	December 1999 18-20 years old	December 2001 20-22 years old	December 2003 22-24 years old	December 2005 24-26 years old
Participation in Secondary Education				
High-school dropper	11 [9] (13)	12 [9] (15)	10 [7] (13)	8 [6] (10)
High-school continuer	13 [11] (15)	2 [2] (2)	1 [1] (1)	0.6 [0.6] (0.5)
High-school graduate	76 [80] (72]	86 [89] (83]	89 [92] (86)	91 [94] (89)
Participation in Postsecondary Education (PSE)				
Never attended PSE	46 [40] (51)	28 [23] (34)	23 [18] (28)	21 [15] (26)
Attended PSE	54 [60] (49)	72 [77] (66)	77 [82] (72)	79 [85] (74)
Attended college/CEGEP[1]	23 [26] (20)	26 [28] (24)	26 [28] (25)	26 [28] (25]
Attended university[2]	21 [24] (19)	33 [37] (30)	37 [41] (34)	40 [44] (36)
Attended other PSE	10 [10] (10)	12 [12] (12)	13 [13] (13)	13 [13] (14)
PSE graduate	8	22	44	60 [62] (57)
PSE graduate continuer	4	8	12	16 [17] (14)
PSE continuer	49	38	20	9 [11] (7)
PSE drop-out	5	10	12	15 [13] (17)

Sources: from YTIS Survey, first panel cycles 1 to 3, Shaienks, Eisl-Culkin and Bussière (2006) and author's calculation for cycle 4; second panel, cycles 1 to 4, Shaienks and Gluszynski (2007). Estimated total number of youth by December 2005: 1,220,000.
Notes: 1. Only. 2. Attended both College/CEGEP and university.

Knighton and Bussière (2006), using results obtained with YTIS-A (PISA)[2] cohort and cycle 3 data of Canadian youth which is slightly younger (age 15 in 2000 and age 19 in 2004) than the YTIS-B cohort compute more precisely high school graduation ages. For this particular cohort, by 2004, 87% have obtained their high school diploma by age 19, 5% were still in high school; and 7% are high-school droppers.

These facts about the dynamics of schooling must be considered when using NLSCY respondents for studying post-secondary attendance. A significant proportion (10%) of the 18 year-olds is made up of high-school continuers (1,372 respondents have obtained their high-school diploma in cycle 6, 986 are PSE participants and for 386 we do not observe their PSE choices (see Table 2). The oldest individuals in cycle 6 are 21 years old and some of them, we assume, will attend PSE institutions later given what we know form YITS-B cohort.

The preceding descriptive studies based on YTIS data sets and the study of Hango and de Broucher (2007) which is both descriptive and econometric (multinomial logistic regression on schooling pathways based on cycle 3) pinpoint the same background factors associated with PSE participation. For demographic factors, data show increases with age and a large gender gap (in favour of women); a higher university participation rate for visible minorities; less participation by youth from rural communities who are more likely to attend a college. For family factors, all the studies identify the following variables: increased parental education is associated with a larger probability of attendance in a university program; living with both parents during high

[2] Programme for International Student Assessment.

school, and value that parents place on education (level of parental expectations according to respondents) have similar effects as well as the mother tongue (English or French compared to other languages). The YTIS for the 18-20 years old cohort does not collect information on parental family income but did collect such information for the 15 years old cohort[3]. Knighton and Bussière (2006) estimate a PSE attendance and a high-school graduation model when they are 19. The results for high school graduation and PSE participation show a negative and significant effect only for the lowest income quartile. For personal characteristics, high school graduation and PSE participation is significantly associated with higher school engagement (positive interactions both academically and socially) and academic performance (good grades) in high school. The results also show that higher levels of reading abilities are associated with a higher rate of high school completion and PSE participation at age 19, after controlling for family and background factors. Finnie and Mueller (2007) use data from the YTIS-A (PISA) cohort in cycle 3 (the youth are aged 19 years in December 2004) and estimate multinomial Logit models of college and university participation for the youth in provinces others than Québec. For family factors, family income (which is available in cycle 1 for this cohort) and parental education remain influential factors of PSE participation, and university in particular. For personal characteristics, a major factor associated with university participation is reading ability (the main element measured by the 2000 PISA test which all the youth in the sample have taken). Overall high school grades are also positively correlated with university participation. High school engagement, self-perception, social support and parent's behaviours (monitoring, nurturance and inconsistent discipline) have some influence depending on gender and PSE level.

Belzil and Hansen (2006) estimate the effects of expected future earnings[4], parental education and individual attributes on educational attainments in Canada, and on the choice of any post-secondary program (college is not separated from university) using data extracted from the 1991 School Leavers Survey (as well as from the follow-up survey in 1995). They estimated a number of specifications of a grade transition model (ordered probit on observed grade levels and a hazard function model of grade transitions[5]); for the type of post-secondary education attended (conditional on attendance), a standard multinomial Logit model is used. Their results for the ordered probit model suggest that parental education is important: regarding mothers' education, schooling beyond high school has a positive and significant effect on their children's education while schooling up to and including high school has no significant effect; for fathers' education however, all categories of education are positive and significant and the magnitudes of the estimated coefficients are substantially larger than those for the mothers' education. For the grade transition model, they find that the father's education influences schooling decisions, but mainly at lower grade levels (high school or less) while the mother's education is generally not significantly related to schooling attainments. Students who report positive attitudes about school and score mainly A's in high school are more likely to attend post-secondary education than those who reported mainly C's. It also appears that a substantial increase in highly educated workers' earnings will significantly affect the high school completion decision as well as the

[3] This cohort youth aged 15 in 2000, also participated in the Program for Student International Assessment (PISA) conducted in 2000 by the OECD. In 2000, reading was the major domain of PISA, and more marginally mathematics and science.

[4] They are measured by average earnings by educational attainments and province from census data. It is worth stressing that they are not able to separately identify effects of expected future earnings from provincial specific effects.

[5] It is defined as high school or less, and more than high-school.

participation decision in post-secondary education. Hansen (2006, 2007) examines returns and labour market outcomes by education level.

Recent American studies in this area downplay the role of credits constraints, student aid or parental income for postsecondary enrolment (Caneiro and Heckman, 2001, 2003; Dynarski, 2003; Keane and Wolpin, 2001).[6] The research literature insists on the long term effects of parental investment, skills accumulation in different life cycle periods and family background characteristics (Todd and Wolpin, 2007; Cunha, 2007; Cunha and Heckman, 2007; Cunha et al., 2006; Heckman, 2007; Ermisch and Francesconi, 2001). On the other hand, life cycle schooling/work decisions and attainments are best modeled in a dynamic setting to take into account selection and heterogeneity problems (Cameron and Heckman, 1998, 2001).

Carneiro and Heckman (2002) and Keane and Wolpin (2001) offer an explanation as to why empirical strategies produce different conclusions with respect to parental income. They argue that higher-income parents have stronger preferences and aptitudes for education, which they transmit to their children. Those parents also tend to invest more and better resources in the development of their children's scholastic abilities from the youngest age. According to this view, the post-secondary enrolment gap between high- and low-income youth originates more from differences in academic ability – the cumulative product of past parental investment − than from differences in access to parental funds at the time of enrolment. This is because low-income youth have not benefited from as much investment in the development of their academic potential, and therefore are less likely to graduate from high school and to attend PSE, irrespective of the level of tuition fees. Therefore, PSE is more a question of qualification than parental income at the moment the child must choose his educational pathway.

[6] Using less sophisticated modelling approach, recent British studies (Blanden and Machin, 2004; Blanden and Gregg, 2004) suggest that family has a 'causal' relationship with educational attainment and access to higher education has not been more equally distributed.

3. Research methodology

The youth aged 18 to 21 in Cycle 6 (the latest wave of data available) of the NLSCY were aged between 8 to 11 in Cycle 1 (1994-95) and they and their family have been surveyed, in principle, every two years. The first panel of Table A1 presents the sample size of the 8-11 years old cohort in cycle 1 and in the following cycles. Unfortunately, the attrition rate of youth between cycles 5 and 6 is very high, in contrast to the preceding cycles. For cycle 5 in 2002-2003, 4,424 youth aged 16 to 19 years were surveyed. But two years later in 2004-2005, only 2,982 youth aged 18 to 21 are retained in the survey.[7] These ages are a turning point in terms of educational choices whether completing or not their high-school diploma or participating or not in postsecondary education. Apart from attrition, the difference in educational systems across provinces (see Figures 1 and 2 for a schematic presentation of differences in pathways) presents an additional problem. Although, the province of Ontario has abolished in 2002-2003 grade 13[8] making the educational system similar to those existing in the other provinces, which could explain low graduation rates in that province in our data there remains the Québec exception (see Figure 2). The college level which is mandatory for admission in Québec's universities renders more complex the analysis of the transition from high school to university.[9] The last panel of Table A2 presents the number of respondents in cycle 6 by age and educational "system" leading to postsecondary education participation in Canada. Given the small sample size in Québec and the role of college in PSE, it is not possible to analyse PSE participation by level for this province alone. For some regressions, the presence of Québec youth in the sample makes empirical sense and a Québec dummy for living in Québec will be added as an explanatory variable in the model. More precisely, we believe that it is warranted to add Québec youth in a model where the dependent variable is PSE attendance and not type of PSE (e.g., college or university).

Research methodology

1) The simplest modeling approach for the investigation of education choices made by teenagers is a utility maximization framework with discrete alternatives. For example, in high-school, teens can choose to drop-out or graduate. At a later stage, if they complete high-school, they can end their schooling, move to a community college, or attend university. Therefore, assuming independence between schooling decisions at different stages, a Logit for the first-decision and a multinomial Logit for the second can be used to determine the probabilities of graduating HS and then the probability of a PSE. At each stage, from the estimated model, we can measure how the inclusion of current and past values of test scores affects the estimate of the current income effect on schooling choices. If the income effect becomes either not significant or very small, this would provide some evidence that selection into post-secondary is in some sense determined much earlier than the year when the choice is actually made. This is the main advantage of the NLSCY compared to other data sets that can model education choices in a longitudinal setting such as YITS-A or the SLID, i.e. we can control for factors occurring ten (or

[7] In the NLSCY, when a youth is 18 years old she/he become the respondent for family information. Family information's are provided by the person most knowledgeable (PMK) of the child when aged less than 18 years. Other information's are collected directly from the questionnaire addressed to the children aged 10 to 17 years.
[8] In cycle 5, Ontarians youth can be observed in grade 13.
[9] For the sample of respondents used in cycle 6, approximately 47 (14) respondents (in Québec) are participating in postsecondary education (college or university) without a high-school diploma.

8, 6, etc) years before schooling decisions are actually made. A drawback of the NLSCY is that choices can be observed only every two year.

2) Estimating post-secondary conditional choices, given graduation from high-school with a Logit or multinomial Logit may lead biased estimates because of self-selection. Indeed, it is reasonable to believe that unobserved factors that determine graduation from high-school will also affect PSE choices, a standard case of selection bias. As shown by Heckman and Cameron (2001), this can be resolved by a Logit specification modified to incorporate some time dependence in the unobserved factors determining schooling choices. They suggest the inclusion of a random effect in the utility function that determines educational choices at each stage. This effect is assumed to take N possible values (determined by estimation). Each value of this effect corresponds to a particular type of individual (e.g. if N=2, high- and low-ability types). A probability which must be estimated is assigned to each type. Our paper will present estimates from this procedure and thus measure the amount of selection bias from static models.

Analytical framework

1. We start by estimating the probability of graduating from high school. From a policy perspective, there is considerable value for prediction.

2 We estimate the probability of attaining the PSE level given a high school diploma. The estimates from this model can be used under certain assumptions as guides for interventions that can increase the probability of a PSE for a high school graduate. A simple OLS approach will first be used to analyse the predictive power of key explanatory variables. OLS possesses the advantage of an easily understandable measure of fit the R2. In the sample, we do observe youth who report a PSE education without graduation from high school but in a very small number so we ignore them in the analysis.

3. The results from predictive equations are helpful identifying children at a young age who are not only at risk of not graduating but also at risk of not pursuing a PSE education even if they graduate. For this purpose, we estimate a multinomial Logit model with the dependent variable categorized in four possible cases: high school drop-outs, high school participants, high school graduates and non PSE participants, and PSE participants.

4. Finally we first examine if self selection matter (because high unobserved ability individuals are over represented for certain groups conditional on high school by estimating a Sequential Logit with un-observed heterogeneity (e.g. we assume two unobserved types of individuals, low- and high-ability) to assess the bias from the simple sequential model. Second we estimate a simple Sequential Logit model and from the estimates, compute the impact of key variables on the overall probability of a PSE which includes the Probability of high school graduation.

4. Survey, data set and variables

Survey

The data used are provided by Statistics Canada's National Longitudinal Survey of Children and Youth (NLSCY) which is a probability survey designed to provide information about children and youth in Canada. The survey covers a comprehensive range of topics including childcare, information on children's physical development, learning and behaviour as well as data on their social environment (family, friends, schools and communities). The NLSCY began in 1994-1995 and data collection occurs biennially. The unit of analysis for the NLSCY is the child or youth. Since the NLSCY objectives are to produce longitudinal and cross-sectional estimates, several populations are targeted.

The NLSCY has a diversity of components which have changed over time and according to the age of the child surveyed. In each NLSCY household, for each selected child/youth, a question is asked to identify who in the household is most knowledgeable about the child. This person is labelled the PMK. The PMK provides the information (child component) for each selected child when he or she is between 0 and 17 years of age in the household and is asked questions about him/her and his/her spouse/partner. The PMK is usually the child's mother (biological or not), but it can also be the father, a step-parent or an adoptive parent who lived in the same dwelling. An adult component was created for the PMK and his/her spouse or partner. Only the PMK or his/her spouse is permitted to answer the questions in this component. A youth component is used for individuals aged 16 years or more. The youth is the only person permitted to answer the questions in this component, whether he/she was living in the family home or not. The list of subjects covered is: moving out of the parental home; education; labour force participation; career aspirations; income; health; activities; social support. In addition, respondents between 10 and 17 years of age completed a self completed questionnaire on various aspects of their lives. The youth was given the questionnaire during the interview and asked to complete it himself/herself. To ensure confidentiality, the child placed the completed questionnaire in an envelope, sealed and gave it to the interviewer. The self-completed questionnaires consisted of a set of five booklets, one for each age group. The subjects covered by each age-group section in the booklet may be (depending on age): friends and family; school; about me; feelings and behaviours; my parent(s); smoking, drinking and drugs; puberty; activities; dating/my relationships; health; work. Youth 18 or older respond for themselves and the PMK provides no information on the family. Therefore, most current values for the family variables are not observed when the child turns 18.

For the purpose of our study (PSE participation), the NLSCY has two main strengths compared to other surveys such as the YTIS or the SLID. First, the survey covers a comprehensive range of topics collected over many waves. The youth aged 18 to 21 years in cycle 6, were 8 to 11 years old in cycle 1. For most of them the child and adult components were collected (in principle) for 5 cycles (when the child is aged less than 18; see Table A1). Depending on the age of the child, the collected measures, scores or scales and social, economic, and demographic features of the household are repeated for a large number of youth. Second, the survey administered tests measuring "skills": a math test in all cycles for children in Grade 2 or above, ranging in age from 7 to 15; a reading test in cycles 2 and 3 for grade school children; other different cognitive tests

or assessments, according to ages for those aged 16 to 21. Other variables relate to non-cognitive skill, family functioning, parenting, and problematic behaviour.

The NLSCY has weaknesses. Although it is not particular to this survey, some of the questions are not answered, in particular questions regarding school by school officials or teachers, resulting in partial non-response (e.g., children attending school at a level higher than kindergarten, the survey in cycles 1 to 4 collected information from questionnaires mailed to teachers and principals on the characteristics of their school and teachers (by the principal's questionnaire[10] and the teacher's questionnaire[11]). The NLSCY does not report their response rates but they are low in particular in cycle 4 (for some variables, the response rate is barely 50%).[12] For the tests such as reading and mathematics, a significant number of respondents have missing scores. But, in our case we will be using the mean value of the test score over all cycles, minimizing the number of missing values. Because math tests are taken more often than reading out math achievement variable may provide a superior measure than our reading achievement variable.[13]

Finally, for the purpose of this study, the sample size of the 18- to 21-year-olds, compared for example to the YTIS cohort's samples in cycles 1 to 4, is small. Moreover, as mentioned earlier, the attrition rate between cycle 5 and cycle 6 was rather large: from 4,424 to 2,982 respondents (see Table A1). Fortunately, cycle 7 of the survey will have retrieved most of these 4,424 respondents, and therefore, a study of this kind will be more conclusive. It will be simple update the data with cycle 7, given the amount of work that was done for this paper.

Data set

In cycle 6 (2004-2005) there are 4,695 longitudinal "youth" respondents which constitute the 16-21 year-old cohort. After excluding 128 respondents from the cohort classified as age unknown (dead or moved to another country) and youth who are 16 or 17 (too young to be in a position to graduate), we are left with 2,982 youth aged 18 to 21 years which constitute the main data set.

Table 2 presents their ages (approximately 25% in each age group), their gender (49% are women and 51% are men) and some of the main educational statistics that we will be focusing on in the empirical part of the paper. The relevant proportions are the following: 10% of the respondents are in high school (HS); 11% are out of school with no HS diploma; 77% are HS graduates. For the HS graduates, 18.7% had never attended PSE at the time of the survey (autumn 2004/spring 2005); 81.3% can be defined as PSE participants. Approximately 2% are PSE participants without a high school diploma. For all provinces except Québec, university is a choice of 53.9% of respondents compared to 46.1% for college.

[10] This questionnaire collected information about the teaching methods used in the school, the availability of educational resources, and the social atmosphere in the school. Hence, the Principal's Questionnaire was about school policies and the educational environment and not about a specific child.

[11] These questionnaires dealt with the child's academic performance and behaviour at school, the teacher's methods of instruction and the atmosphere in the classroom.

[12] For the tests, consent had to be given by the parent(s) and the school boards. For cycle 1, the NLSCY received results for about 50% of eligible children; for cycle 2, the percentage was 74% and for cycle 3, it was 54%.

[13] See Lefebvre et al. (2008) who exploit the longitudinal nature of the school collection information and math test scores.

Table 2: Sample size, educational status and postsecondary education participation (PSE) by region, youth 18-21 year olds in cycle 6 of the NLSCY, 2004-2005

	Canada	Québec	Other Provinces	Women	Men
Observations	2,982	531	1,721	1,591	1,391
Weighted observations[2]	1,509,944	349,943	593,662	736,215	773,729
Age					
18	815	137	460	407	408
	24.8%	25.3%	24.6%	24.7%	24.9%
19	758	140	440	406	352
	25.5%	25.0%	25.4%	25.6%	25.4%
20	746	137	426	412	334
	25.1%	24.6%	25.0%	25.1%	25.1%
21	663	117	395	366	297
	24.6%	25.0%	25.0%	24.6%	24.6%
Educational status					
High-school continuers	310	51	259	132	178
	10.2%	9.4%	10.4%	8.5%	11.8%
High-school droppers	299	71	228	129	170
	11.2%	15.7%	9.6%	9.5%	12.5%
High-school graduates	2,326	395	1,931	1,315	1,011
	77.2%	72.7%	78.6%	80.9	73.7
High-school graduates non attending PSE	505	26	479	220	285
	18.7%	9.7%	21.3%	13.9%	23.7%
High-school graduates attending PSE	1,821	369	1,452	1,095	726
	81.3%	90.3%	78.7%	86.1%	76.3%
High-school droppers attending PSE	47	14	33	15	32
	1.9%	3.0%	1.6%	5.9%	7.5%
PSE participation by level					
All	1,868	383	1,485	1,110	758
	64.3%	72.3%	63.3%	70.8%	58.2%
University	885	93	792	575	310
	30.5%	18,5%	34.1%	35.8%	25.4%
College	983	290	693	535	448
	33.9%	49.4%	29.2%	35.0%	32.8%
Proportion attending PSE by level					
University	47.3%	27.2%	53.9%	50.5%	43.6%
College	52.7%	72.8%	46.1%	49.5%	56.4%

Sources: author's calculation from the NLSCY's cycles 4 to 6.

Notes: 1. Post-secondary education (PSE) is defined as any type of schooling higher than high school. 2. Weighted observations have been used (cycle 6 longitudinal weights) to calculate the percentages presented in this table. Other provinces: all provinces less Québec.

In Québec most PSE participants are observed in college which is the usual path to accede university. Québec youth have a much higher rate of dropping out of high school (i.e., observed out of school with no HS degree at the time of the cycle 6 survey), 16.9%. There are large gender

gaps. Women are more likely to graduate from high school and at an earlier age (there are less high-school continuers by age 18 or more). They participate more in PSE as high-school graduates (10% gap compared to men) and more access university rather than college (10% gap compared to men).

Variables available for the regression analysis[14]

In preliminary estimations for variables that have different values over the cycles, we used for the estimations the value for the cycle 1. We did this to see whether the child's situation 10 years before making schooling choices has a strong impact, if yes, this justifies modes of intervention at an early age. Also, some variables are particular to the age of the child. For example, some scales are for 4 to 11 year olds, while some are for the 10 to 15 year olds. Since youth are not respondents in all cycles, we finally used most of the times the earlier values of the variables or their average when more than one value was available over the cycles of the survey. Table A2 presents these variables by domain and their cycle of ages/availability and those that we used in the final estimations. The variables are:

1. Family background-parent(s) reported: province of residence; education levels of the PMK; family type (two biological parents, two parents, one parent/other guardian; change in family type before age 18);[15] siblings (zero, one or two or more); community size (rural, less than 100,000 inhabitants, 100,000 inhabitants or more); PMK not born in Canada; family income in 2002 dollars averaged over the cycles when family income is observed (before age 18); family income quartiles.

2. Community-parent(s) reported: The NLSCY collects information[16] on the community such as a neighbourhood safety score (the score varies between 0 and 15, a high score indicating a low degree of neighbourhood safety), neighbours score (the score varies between 0 and 15, a high score indicating a high degree neighbour cohesiveness) and a neighbourhood problems score (the score varies between 0 and 15, a high score indicating a high degree of neighbourhood problems). From cycle 4, a neighbourhood structure score is calculated using these questions and asked to the 16-17 year olds (a high score indicates a high degree of neighbourhood structure and a low score indicates a low degree of neighbourhood structure). Except for community size, these variables were never significant in our regressions. The estimations results are not presented but available from the authors.

3. Youth personal characteristics and objective tests: age at survey; health (excellent or very good versus good, fair and poor); normalized[17] averaged math score before 16,[18] normalized

[14] Most of these variables were constructed as a panel data set for cycles 1 to 6 but not necessarily used for the estimation results presented below. An appendix presents the STATA do file developed which can be updated for further cycles of data.

[15] As they are older more youth are classified as "autonomous".

[16] The entire Neighbourhood section was not asked of survey participants in Cycle 2. In Cycle 3, the Neighbourhood section was reintroduced without the safety and problems questions. These questions applied only to families with a child aged less than 16 years.

[17] The standardized score in the file (which takes into account grade level) is transformed by the following operation: the mean of the score is subtracted and divided by the standard error. The normalized score has a mean of zero and a range between -2 and 2.

averaged reading score in cycles 2 and 3;[19] normalized and averaged math score quartiles; normalized and averaged reading score quartile; average value of the cognitive measure (the test covered reading and mathematics) for 16-17 year-olds (cycles 4 and 5),[20] the scores presented on the data file have a mean of 0 and a standard deviation of 1;[21] literacy assessment (cycle 6) for 18-19-year olds;[22] numeracy assessment (cycle 6) for 20-21-year olds.[23] Literacy and numeracy scores are most likely endogenous "skills" correlated with schooling participation. Since they are taken so late in terms of the age of the child, they could depend on the level of schooling attained, causing a problem of reverse causality. We therefore do not use them as regressors in our analyses.

4. *Child or parenting scores on parent-reported scales* (in parentheses, age of child for which measures are taken): depression rating (PMK; 0-15 years); family functioning score (PMK or spouse; 0-15 years); social support (PMK or spouse; 0-15 years); home responsibilities (PMK; 10-13 years); hyperactivity-inattention (PMK; 4-11 years); pro-social behaviour (PMK; 4-11 years); emotional disorder-anxiety (PMK; 4-11 years); physical aggression-conduct disorder (PMK; 4-11 years); indirect aggression (PMK; 4-11 years); property offences (PMK; 8-11 years); health status of the child on a scale of five (from excellent to poor). The NLSCY also includes parenting scales: positive reaction (PMK; 2-11 years); ineffective parenting style (PMK; 2-11 years); consistent parenting style (PMK; 2-11 years); rational parenting style (PMK; 2-11 years); conflicts tactics/resolution (PMK, 12-15 years); parent-child cohesion (PMK, 12-15 years).

5. *Scores on youth-reported scales* (in parenthesis, age of child for which measures are taken): emotional/social capacity quotient, five composite scales and two aggregated measured (10-17

[18] The mathematics test (computation exercise) is an objective indicator of the child's academic performance in mathematics. It was administered to children in Grade 2 or above, ranging in age from 7 to 15. It consisted of a set of nine booklets of varying levels of complexity. The level was determined by the child's grade. Since cycle 4, the test is administered by the interviewer in the child's home.

[19] In Cycles 2 and 3, there was both a Math and a Reading test. In Cycle 4, only the math test was administered. The reading test was removed because of time constraints. It was decided that only one test could be administered. The math test was chosen as it has been administered in all previous cycles.

[20] This test comprises questions (items) from the pilot of the Programme for International Student Assessment (PISA 2000) mathematics test that were not in the final version of the PISA test. Since the PISA test was designed for 15-year-olds, the theoretical accuracy of the NLSCY Cognitive Measure test in estimating the ability of advanced respondents was not ideal. However, some NLSCY respondents noted that the test was quite difficult for them, and the test seems to be at least as effective as the NLSCY mathematics tests from previous cycles. There were actually two tests, one for higher-ability youth and a slightly easier one for lower-ability youth. Data from previous cycles was used to pre -select the respondents into the high-ability group and the low-ability group. Each test contained 18 items. Ten items were shared between the two levels of tests. Eight items were unique to the test intended for the high-ability group and eight items were unique to the test designed for the low-ability group.

[21] To obtain the Cognitive Measure score, the three-parameter model from Item Response Theory (IRT) was used.. This scale is standard in IRT.

[22] The Literacy Assessment for 18- and 19-year-olds consisted of 36 questions with an emphasis on extracting information from texts, tables and graphs. The test required a personal visit while the youth component could be completed by phone.

[23] The Numeracy Assessment for 20- and 21-year-olds consisted of 32 questions. It aims to test the ability of young adults to function in society and manage mathematical demands in diverse situations. The test required a personal visit while the youth component could be completed by phone.

years, cycles 5 and 6; and 20-21, cycle 6);[24] general-self (10-19 years);[25] hyperactivity-inattention (10-15 years); pro-social behaviour (10-15 years); emotional disorder-anxiety (10-15 years); physical aggression-conduct disorder (10-15 years); indirect aggression(10-15 years); property offences (10-15 years); parental nurturance (10-15 years); parental rejection (10-15 years); parental monitoring (10-15 years). For the older youth (self-completed): neighbourhood structure (16-17 years); depression (16-17 years); conflict resolution mother/father (16-17 years); friends (10-17 years).

6. School "quality" index from principals and teachers questionnaires

The NLSCY collects information on the school characteristics of children in elementary and secondary levels schools. We use an index of school "quality" constructed by Lefebvre et al. (2008a) as an explanatory variable in our regressions. The score ranges between 0 and 10, higher scores reflecting better schools.

Variables used and descriptive statistics

Table 3 presents our preferred independent variables that were used in all the estimations presented below and their definition. These variables capture family background characteristics (community size, provinces, PMK levels of education, family structure and their changes over time, presence of siblings and "permanent" family income); youth personal characteristics and assessments (age, gender, health status, math and reading scores). We built our specifications on the basis of experimentation and the previous literature on schooling attainment. Several of the potential explanatory variables were either non-significant or produced signs that were clearly counter-intuitive and were not included in the final specification.

Heckman et al. (2006) (and Cunha and Heckman (2006)) in their analysis of the effect of cognitive and non-cognitive abilities on labour market outcomes and social behaviour retain math and reading scores as their preferred measures of cognitive skills. For their measure of non-cognitive skills they use standardized sum of scores on scales of "Locus of control" and 'Self-Esteem". The emotional/social capacity quotient (ages 10 to 17 in cycles 5 and 6; and ages 20 and21 in cycle 6), and the general-self or self-esteem measure (10 to 19 years) correspond to their measures of non cognitive skills. Unfortunately, not all respondents have a scale for the emotional quotient (i.e., had the opportunity to complete these questions from the self-completed questionnaire). Estimations for a sample of respondents with such a measure as a regressor show that the effect of the variable was never significant. The self-esteem measure (few respondents have a missing value) remains the only non-cognitive ability variable in our regressions.

The NLSCY offers a diversity of behavioural scales for both the family/parent(s) and the child/youth. After experimenting with most of them (see Table A2 for the list), the results show that several were insignificant predictors of later PSE participation. Duncan et al. (2006) using

[24] The emotional quotient measures: intrapersonal competencies - self-awareness and self-expression; interpersonal competencies - social awareness and interpersonal relationship; stress management competencies - emotional management and regulation; adaptability competencies - change management; general mood - self-motivation.

[25] The objective of the General-self Scale is to measure the child's overall self-esteem. The self-esteem scale was expanded each year to include the oldest cohort. This means that by Cycle 5 the items making up this scale are asked of all youth aged 10 to 19 years.

six longitudinal data sets to study school readiness and later achievement find that early reading and math skills are the strongest predictors.

Table 4 presents descriptive statistics (percentage distribution or value) of the variables used and of for some supplementary behavioural variables used in preliminary analysis, for all youth by gender and by their educational status observed in cycle 6 and for females and males. The Table shows that very few youth changed of province of residence between cycles 1 and 6, although the proportions for Alberta suggest that some non PSE participant youth may have moved to this province. Overall 18% of youth have a low-education PMK; the proportion is much higher for youth who are high school dropper and very low for youth having accessed to university. Living with in a biological two-parent family (in cycles 1 to 4) is correlated with education attainment. For youth's university participant, the mean level of family's permanent income (cycles 1 to 4) is almost twice the level of those youth who are high school droppers. Youth, having very good scores in math and reading, are more likely to be PSE participant, and in particular at the university level. Curiously, the mean health status reported by the PMK (in cycles 1 to 4) for the child is much higher than the one reported by the youth themselves when they are aged 18 years or more; and, in particular males view themselves in better health than females. There are not much differences between mean of non cognitive measures (such as self-esteem and emotional quotient) and of behavioural scales (parenting or reported by the youth) by educational level for females and males. The differences are larger between females and females pour a given educational status.

Table 3: Definition of controls variables (reference category in parenthesis) and dependent variables

Controls variables	Definition
Woman (man)	Gender
(Age 18)	Age in cycle 6
Age19	Age in cycle 6 Takes the value of 1 if the individual is 19 in cycle 6, 0 otherwise
Age 20	Age in cycle 6 Takes the value of 1 if the individual is 20 in cycle 6, 0 otherwise
Age 21	Age in cycle 6 Takes the value of 1 if the individual is 20 in cycle 6, 0 otherwise
(Atlantic provinces)	Newfoundland, PEI, Nova Scotia, New Brunswick in cycle 6
Que	Takes the value of 1 if the individual resides in Québec in cycle 6, 0 otherwise
Ont	Takes the value of 1 if the individual resides in Ontario in cycle 6, 0 otherwise
Prai	Takes the value of 1 if the individual resides in Saskatchewan or Manitoba in cycle 6, 0 otherwise
Alb	Takes the value of 1 if the individual resides in Alberta in cycle 6, 0 otherwise
BC	Takes the value of 1 if the individual resides in British Columbia in cycle 6, 0 otherwise
(PmkEduHS)	Education level of the person most knowledgeable (PMK) about this child: is high school diploma reported in cycle 1
PmkEduPrim	Takes the value of 1 if PMK highest education is primary in cycle 1, 0 otherwise
PmkEduSec	Takes the value of 1 if PMK highest education is some high school studies in cycle 1, 0 otherwise
PmkEduSomePSE	Takes the value of 1 if PMK highest education is some postsecondary studies in cycle 1, 0 otherwise
PmkEduCollege	Takes the value of 1 if PMK highest education is a college diploma in cycle 1, 0 otherwise
PmkEduUniversity	Takes the value of 1 if PMK highest education is university degree or more in cycle 1, 0 otherwise
(One-parent/guardian)	The child is living with a one-parent/guardian in cycle 1
TwoBioParents	Takes the value of 1 if the child is living with two biological parents in cycle 1, 0 otherwise
TwoParents	Takes the value of 1 if the child is living with two parents in cycle 1 but one or two of them are not biological, 0 otherwise
Separation	Takes the value of 1 if the child experienced a move from a two-parent family to a single parent family during the first 4 cycles, 0 otherwise
PmkImmig	Takes the value of 1 if the PMK in cycle 1 is not born in Canada, 0 otherwise
(Sibling 0)	No sibling in cycle 1
Sibling1	Takes the value of 1 if the child has one sibling in cycle 1, 0 otherwise
Siblings2	Takes the value of 1 if the child has two sibling or more in cycle 1, 0 otherwise
(ComSize1)	The child lives in a town with less 100,000 inhabitants in cycle 1, 0 otherwise
ComSizeRural	Takes the value of 1 if a child lives in a rural community in cycle 1, 0 otherwise
ComSize2	Takes the value of 1 if a child lives in town with 99 999 inhabitants or more in cycle 1, 0 otherwise
SchoolQIndex	Mean value of the school quality variable taken over cycles 1 to 4
(Q1FamInc)	First quartile of mean family incomes of the distribution of mean family income over cycles 1 to 4, $2002
Q2FamInc	Takes the value of 1 if the mean family income over cycles 1 to 4 is in the second quartile, 0 otherwise
Q3FamInc	Takes the value of 1 if the mean family income over cycles 1 to 4 is in the third quartile, 0 otherwise
Q4FamInc	Takes the value of 1 if the mean family income over cycles 1 to 4 is in the fourth quartile, 0 otherwise

Table 3 end: Definition of controls variables (reference category in parenthesis) and dependent variables

(Q1Math)	First quartile of mean math scores of the distribution of mean math scores income over cycles 1 to 4
Q2Math	Takes the value of 1 if the mean math scores over cycles 1 to 4 is in the second quartile, 0 otherwise
Q3Math	Takes the value of 1 if the mean math scores over cycles 1 to 4 is in the third quartile, 0 otherwise
Q4Math	Takes the value of 1 if the mean math scores over cycles 1 to 4 is in the fourth quartile, 0 otherwise
(Q1Reading)	First quartile of mean reading scores of the distribution of mean reading scores over cycles 2 and 3
Q2Reading	Takes the value of 1 if the mean reading scores over cycles 2 and 3 is in the second quartile, 0 otherwise
Q3Reading	Takes the value of 1 if the mean reading scores over cycles 2 and 3 is in the third quartile, 0 otherwise
Q4reading	Takes the value of 1 if the mean reading scores over cycles 2 and 3 is in the fourth quartile, 0 otherwise
HealthE&VG	Takes the value of 1 if mean health status is Excellent or very good over cycles 5 and 6, 0 otherwise
SelfEsteem	Mean general self score for children aged 10 to 19 years
HyperActInat	Mean value of hyperactivity/inattention score (for children who are less than 11 in cycles 1 and 2
Aggression11/15	Mean value of conduct disorder-physical aggression scores: boys aged 8 to 11 years; girls aged 10 to 15 years
Nurturance	Mean parental nurturance score for cycles 3 and 4 for children aged 10 to 15 years
Dependent variables for educational outcomes	
Educational outcomes	Canada less Québec, 3 outcomes: 1. HSG; 2. CP; 3. UP
Educational outcomes	Canada, 4 outcomes: 1. HSDO; 2. HSC; 3. HSG; 4. CP and UP

Notes: HSG=high school graduate; CP=college participation; UP=university participation; HSDO: high school drop-out; HSC: high school continuer.

Regarding scales related to the family environment, only the nurturance scale showed promise while others were far from significant. A problem with these scales is that they are highly correlated so that when they are all included in the same regression, standard multi-collinearity problems emerge (e.g. unexpected sign for certain coefficients, large standard errors). Future research using factor analysis, following Cunha et al, reducing the number of explanatory variables in models of PSE attendance is a promising avenue to clarify the role of the family in child development. We retained parental nurturance which was well behaved and produced some significant effects and other coefficients reacted strongly to its presence in the regressions.

Socio-emotional behaviours, except hyperactivity-attention and aggression, were generally insignificant predictors of later academic performance even among children with relatively high levels of problem behaviour and for children from high and low socio-economic backgrounds. Hyperactivity and aggression are variables in the psychology literature which are given an important role in child development, particularly in the works of Richard E. Tremblay and his co-authors (Tremblay et al., 2008; Fontaine et al., 2008; Leblanc et al., 2008). Therefore, we included a measure for each of these behaviours in the analysis.

Table 4: Descriptive statistics by gender and schooling status by cycle 6, 18 to 21 year olds, NLSCY Canada 2004-2005 (continued on next page)

Variables	All youth		High school dropper		High school continuer	
	Women	Men	Women	Men	Women	Men
Age 18	25	25	16	16	72	57
Age19	25	26	40	40	14	20
Age 20	25	25	20	20	10	15
Age 21	25	25	24	24	5	9
Atl C1/C6	9/9	8/8	6/7	7/6	9/11	7/7
Que C1/C6	23/24	23/23	29/29	36/36	18/18	23/23
Ont C1/C6	37/37	38/38	39/36	30/30	49/48	47/47
Prai C1/C6	8/7	8/8	7/6	12/12	5/5	7/6
Alb C1/C6	11/12	10/12	17/18	7/9	10/10	6/7
BC C1/C6	12/11	12/12	3/3	8/8	87	10/10
PmkEduPrim	4	4	15	16	4	2
PmkEduSec	14	14	17	27	12	10
PmkEduSecHS	19	18	19	17	22	22
PmkEduSomePSE	31	26	31	22	44	29
PmkEduCol	17	22	16	13	11	27
PmkEduUni	14	16	2	5	6	9
TwoBioParents	74	77	44	58	68	63
TwoParents	10	9	16	11	11	11
Singleparent/guardian	16	14	39	31	21	26
Separation	15	16	18	21	12	13
PmkImmig	16	18	12	17	11	22
Sibling0	13	13	23	7	12	19
Sibling1	45	45	35	50	51	45
Siblings2	42	42	42	42	37	36
ComSizeRural	20	21	25	30	16	19
Com Size1	23	21	22	22	27	20
ComSize2	57	58	53	47	57	61
MSchoolQIndex	3.7	3.7	3.4	3.4	3.6	3.7
MFamInc ($2002)	73,928	70,904	44,106	48,369	62,467	67,034
Q1FamInc	26	24	57	54	25	23
Q2FamInc	25	27	23	18	21	33
Q3FamInc	24	24	10	19	35	24
Q4FamInc	25	25	10	9	19	19
Q1Math	19	17	33	34	27	22
Q2Math	30	31	32	33	35	42
Q3Math	28	29	21	26	23	25
Q4Math	24	24	14	8	15	10
Q1Reading	21	27	35	41	31	35
Q2Reading	21	24	16	21	17	32
Q3Reading	29	25	20	27	34	21
Q4reading	28	24	29	10	18	11
MHealthE&VG_C1-C4	88	85	76	76	81	83
MHealthE&VG_C5-6	60	68	68	56	40	67
MSelfEsteem	12.6	13.0	11.5	12.4	12.2	13.1
MHyperActInat	4.3	5.4	5.3	8.1	6.8	7.1
MAggression11/15	1.0/0.8	1.4/1.3	1.3/1.9	2.5/1.7	1.4/1.2	1.9/2.3
MNurturance	20.4	19.8	17.2	18.4	21.0	20.1
MProsocial15	13.7	11.1	12.3	10.2	14.0	11.2
MFamilyFunctioning	8.4	8.5	9.3	9.3	8.1	9.0
MMonitoring	14.9	14.0	13.7	13.6	14.9	14.5
MInefficientparenting	8.7	9.0	9.0	10.8	8.9	9.5
MConsistentparenting	14.9	15.1	14.5	13.8	14.8	14.6
MEmotional quotient	29.3	28.9	26.6	26.9	26.7	26.3
Observations	736,215	773,729	69,891	96,592	62,272	91,350

Table 4 end

Variables	High school graduate no PSE		College participant		University participant	
	Women	Men	Women	Men	Women	Men
Age 18	34	30	21	15	17	19
Age19	31	26	21	27	27	26
Age 20	23	26	29	25	28	29
Age 21	13	18	29	33	28	26
Atl C1/C6	12/11	12/10	5/5	7/6	12/12	9/9
Que C1/C6	10/10	13/12	35/36	32/31	16/15	14/12
Ont C1/C6	35/34	32/31	32/31	30/30	40/42	52/53
Prai C1/C6	13/12	10/10	6/5	5/5	9/8	8/8
Alb C1/C6	15/18	15/19	10/13	12/13	9/8	9/8
BC C1/C6	15/15	19/17	11/10	15/15	15/15	8/10
PmkEduPrim	7	4	3	1	1	0
PmkEduSec	14	23	21	14	7	5
PmkEduSecHS	26	16	21	18	16	18
PmkEduSomePSE	29	28	30	28	29	21
PmkEduCol	17	20	17	26	19	21
PmkEduUni	7	9	8	12	27	35
TwoBioParents	76	77	75	84	81	86
TwoParents	5	11	11	8	9	6
Singleparent/guardian	19	13	14	9	10	8
Separation	17	25	18	20	22	6
PmkImmig	18	14	12	17	22	22
Sibling0	9	13	13	15	11	10
Sibling1	51	39	46	45	45	49
Siblings2	40	47	41	41	45	41
ComSizeRural	24	28	23	25	16	9
Com Size1	25	25	24	22	21	17
ComSize2	52	47	52	54	63	74
MSchoolQIndex	3.8	3.6	3.5	3.8	4.0	3.7
MFamInc ($2002)	57,395	63,097	67,168	69,301	91,322	91,233
Q1FamInc	34	28	27	22	14	10
Q2FamInc	30	27	29	26	21	29
Q3FamInc	2	27	24	24	27	23
Q4FamInc	14	18	20	27	38	38
Q1Math	27	19	20	15	9	7
Q2Math	39	37	31	31	24	19
Q3Math	27	29	29	27	30	34
Q4Math	6	15	20	28	37	40
Q1Reading	36	29	24	28	9	13
Q2Reading	19	21	28	28	18	17
Q3Reading	28	29	27	20	34	30
Q4reading	17	21	22	23	39	39
MHealthE&VG_C1-C4	87	85	90	88	92	88
MHealthE&VG_C5-6	55	65	56	69	76	78
MSelfEsteem	12.5	12.7	12.5	13.3	13.1	13.2
MHyperActInat	4.5	5.7	4.5	5.1	3.0	3.6
MAggression11/15	1.1/1.0	1.5/1.4	1.1/0.7	1.2/1.1	0.8/0.3	0.9/1.0
MNurturance	20.1	19.2	20.1	20.4	21.5	19.9
MProsocial15	13.0	10.7	13.7	11.0	14.3	11.8
MFamilyFunctioning	8.9	9.6	8.4	8.2	8.1	7.6
MMonitoring	13.9	13.6	14.7	14.0	14.6	14.3
MInefficientparenting	8.7	9.4	8.8	8.5	8.6	8.1
MConsistentparenting	15.5	15.4	14.6	15.4	15.2	15.1
MEmotional quotient	28.9	27.6	29.4	29.8	30.5	30.1
Observations	83,043	135,432	257,763	254,546	263,246	196,554

Notes: author's calculation from cycles 1 to 6 weighted data and rounding of percentage. See Tables 3 and A3 for definition of variables. For some variables, percentage is calculated only for responses. M: mean over cycles of availability.

5. Econometrics results

All estimations, were done with STATA10, are weighted by cycle 6 longitudinal weights, and use the bootstrap weights provided by Statistics Canada for variance estimation, unless specified otherwise.[26] Therefore, the results are representative of the longitudinal 1994 cohort, where immigrants are under-represented compared to a cross-sectional cohort of 18 to 21 year olds in 2004. Preliminary estimations for the overall sample of women and men always show a large gender effect. Since a diversity of educational status measures for women and men (see Table 1) indicate large gender gaps, all estimations are conducted separately by gender to capture the main elements at the origin of diverging schooling achievements.

5.1 Ordinary least squares results

Our opening econometric strategy is simple. We estimate for young men and women including Québec's respondents the probability of: a) obtaining a high-school diploma; b) PSE participation conditional on being a high-school graduate (and not type of PSE attendance, i.e., college or university).[27] The youth are aged 18 to 21 from all provinces, and observed in cycle 6 of the NLSCY. The estimations for women are in Tables 5 and 7 and for men are in Tables 6 and 8. Estimations are conducted separately for Québec respondents and results are presented in Tables A3-A4. We do this because the PSE education system is different and almost all high school graduates participating in PSE are observed in college.

These series of regressions are performed under the assumption that the probabilities estimated are linear. Although the problems with such a model are well-known,[28] this strategy will serve us well given our preliminary objectives. One of the advantages of working with the linear probability model is that a R2 measure of fit is easily computed and therefore permits an evaluation of the explanatory power of different variables included as regressors.

Our modeling will highlight explanatory variables that are in the NLSCY and not in other data sets, in particular math and reading scores from tests performed in grade or high school, scores of hyperactivity and aggression, quality of schools, and parental nurturance. These are particularly interesting because they can inform about the probability of attending a PSE when the child is particularly young. Our preferred specification includes as explanatory variables, the age of the child, the province of residence at the moment of making the choice of attending a PSE, the PMK's education, family type, whether the child experienced a separation between cycles 1 and 4, the number of siblings, PMK is immigrant, the size of the community, the average quality of the school the child attended between cycles 1 and 4, "permanent" family income (mean income measured over four cycles), measures of achievement in reading and mathematics, and finally measures of health, self-esteem, hyperactivity-inattention at ages 8-11, self-esteem, and

[26] The standard errors of the Sequential Logit, with uncorrelated unobservable error terms, were not computed using the bootstrap weights as the STATA procedure could not implement it; however the regressions are performed with weights. The Sequential Logit with a random effect was performed without weights. Since this was done for comparison with the Sequential Logit with no correlation, to ascertain bias, the latter was done without weights as well. No policy analysis is done on the basis of these results.

[27] This condition excludes 47 respondents not having this diploma but participating in PSE (42 college and 5 university students) and 310 high-school continuers, a majority of males.

[28] The predicted probabilities obtained from a linear probability model can be higher than 1 or smaller than 0.

aggression (conduct disorders and physical aggression) at ages 8-11 for males and 10-15 for females, and finally, parental nurturance (see text above and Tables 3 and A2 for definition).

The first specification excludes the hyperactivity and aggressions measures as well as nurturance. The second one adds the hyperactivity and aggression measures and the third one includes mean parental nurturance scale measured between ages 10 to 15 years. Adding the behavioural scale variables to the other variables considerably reduces the sample sizes.

Probability of obtaining a high school diploma

The age effects are larger at ages 20 for women and at 21 for men showing that women obtain their diploma on average earlier. The Québec effect is very large, negative and significant for men in this province, reflecting its high drop-out rate, but insignificant for women. A recent document from the Ministry of Education in Québec reveals that in 2007, only 52.1% of males graduated from high-school in 5 years (5 being the number of grades at the HS level) and 65.7% for females, a dramatic gap.

Having a very low-education parent (PMK) reduces significantly the probability of graduation for both women and men. The effect of a high-education parent (college or university degree) is very strong and significant for females but less important (non significant) for males. Living in an "intact" two-parent biological family increases, significantly and substantially, the probability of graduation. The impact for males is almost double (.20) that of females (.11). Having a parent not born in Canada is in general not significant except specification (3) ($p<.01$) for the sample of females (.08). The presence of one or two siblings or more has significant and large impacts according to gender: positive for females and negative for males (except in the final specification which takes into account the nurturing scale), an intriguing result. The impact of the school index of quality is almost never significant (except in one specification for females).[29] Family income quartiles do not affect the probability of graduation except in one specification for the females (Q4, specification (1), $p<.01$).

Despite the fact that the effects of income quartiles are not significant, the ceteris paribus difference in the probability of graduation between Q1 and Q2 is relatively large for females (between .06 and .08) but then is negligible between Q2 and Q3 and Q3 and Q4. For males, the effect is more linear and the difference between Q1 and Q4 is almost .08 for specification (3).

[29] To reduce the number of missing values we average the values. This somewhat blurred the impact of the variable. The index takes a value between 0 and 10; a low value indicates a low "quality" school and a high index a "high" quality index. In preliminary estimations the index was used by cycle. The first index (for cycle1) refers mostly to children in elementary school (1,359 respondents aged 8 to 11) and is not significant but the explanatory power (R2) increases significantly. In cycle 2, the one with the least missing values (1,487 respondents), the index refers to children in the last years in primary school and in their first years in secondary school. It is also not significant. The cycle 3 index is not significant (1,349 respondents). When two or three indexes are used, the cycle 2 index was always significant; but the number of respondents was much smaller (1,142; 878). Results are different for women and men, the cycle 2 index being most of the time significant.

Table 5: OLS estimation of obtaining high-school diploma, 18 to 21 year olds women, NLSCY Canada 2004-2005

Variables	Specifications					
	(1)		(2)		(3)	
Age19	**0.125	(0.049)	**0.106	(0.049)	**0.103	(0.050)
Age 20	***0.284	(0.041)	***0.240	(0.040)	***0.235	(0.041)
Age 21	***0.234	(0.044)	***0.204	(0.045)	***0.209	(0.047)
Que	0.010	(0.057)	-0.017	(0.058)	-0.021	(0.060)
Ont	-0.068	(0.052)	-0.085	(0.053)	*-0.089	(0.053)
Prai	0.045	(0.053)	0.035	(0.052)	0.042	(0.053)
Alb	-0.080	(0.061)	-0.034	(0.058)	-0.033	(0.060)
BC	0.018	(0.059)	-0.023	(0.061)	-0.037	(0.062)
PmkEduPrim	**-0.273	(0.136)	*-0.277	(0.146)	*-0.287	(0.155)
PmkEduSec	0.028	(0.048)	0.064	(0.049)	0.052	(0.054)
PmkEduSomePSE	-0.010	(0.040)	0.006	(0.041)	0.012	(0.043)
PmkEduCol	*0.074	(0.044)	*0.080	(0.045)	**0.097	(0.046)
PmkEduUni	***0.133	(0.043)	***0.123	(0.044)	***0.126	(0.045)
TwoBioParents	*0.097	(0.053)	*0.103	(0.055)	*0.108	(0.058)
TwoParents	0.010	(0.075)	0.047	(0.074)	0.051	(0.078)
Separation	0.011	(0.040)	0.021	(0.041)	0.027	(0.045)
PmkImmig	0.072	(0.045)	0.073	(0.047)	*0.083	(0.047)
Sibling1	*0.089	(0.051)	*0.095	(0.054)	*0.104	(0.054)
Siblings2	*0.096	(0.055)	*0.097	(0.058)	*0.099	(0.057)
ComSizeRural	0.003	(0.036)	0.008	(0.037)	0.016	(0.039)
ComSize2	-0.010	(0.034)	-0.025	(0.035)	-0.021	(0.036)
SchoolQIndex	*0.018	(0.011)	0.017	(0.011)	0.017	(0.011)
Q2FamInc	0.057	(0.048)	0.078	(0.051)	0.081	(0.053)
Q3FamInc	0.078	(0.051)	*0.092	(0.054)	0.089	(0.056)
Q4FamInc	0.068	(0.056)	0.072	(0.061)	0.077	(0.062)
Q2Math	0.080	(0.051)	0.059	(0.054)	0.044	(0.057)
Q3Math	0.073	(0.049)	0.064	(0.049)	0.044	(0.052)
Q4Math	*0.092	(0.055)	0.071	(0.057)	0.051	(0.060)
Q2Reading	0.080	(0.053)	0.065	(0.053)	0.069	(0.058)
Q3Reading	0.063	(0.048)	0.040	(0.049)	0.044	(0.052)
Q4reading	0.052	(0.052)	0.038	(0.051)	0.049	(0.056)
HealthE&VG	***0.086	(0.032)	**0.081	(0.035)	**0.084	(0.037)
SelfEsteem	**0.017	(0.007)	0.006	(0.007)	0.004	(0.008)
HyperActInat	-	-	***-0.014	(0.005)	**-0.012	(0.006)
Aggression15	-	-	***-0.045	(0.012)	**-0.043	(0.013)
Nurturance	-	-	-	-	0.003	(0.004)
Constant	-0.011	(0.128)	*0.252	(0.131)	0.209	(0.143)
Observations	1,342		1,230		1,168	
Adj. R^2	0.211		0.253		0.247	

Notes: 1. Bootstrapped standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: OLS estimation of obtaining high-school diploma, 18 to 21 year olds men, NLSCY, Canada 2004-2005

Variables	Specifications					
	(1)		(2)		(3)	
Age19	***0.184	(0.049)	***0.174	(0.048)	***0.167	(0.049)
Age 20	***0.228	(0.048)	***0.208	(0.046)	***0.206	(0.046)
Age 21	***0.264	(0.045)	***0.216	(0.043)	***0.212	(0.045)
Que	***-0.259	(0.056)	***-0.200	(0.051)	***-0.211	(0.055)
Ont	-0.069	(0.045)	-0.063	(0.043)	**-0.094	(0.041)
Prai	-0.061	(0.050)	-0.018	(0.048)	-0.048	(0.050)
Alb	-0.014	(0.050)	0.015	(0.049)	-0.033	(0.046)
BC	-0.024	(0.063)	-0.022	(0.063)	-0.009	(0.057)
PmkEduPrim	**-0.308	(0.135)	*-0.223	(0.121)	-0.080	(0.115)
PmkEduSec	0.013	(0.067)	0.016	(0.063)	0.004	(0.067)
PmkEduSomePSE	0.028	(0.047)	0.032	(0.046)	0.033	(0.049)
PmkEduCol	0.065	(0.046)	0.073	(0.045)	0.017	(0.046)
PmkEduUni	0.048	(0.053)	0.065	(0.052)	0.028	(0.053)
TwoBioParents	***0.233	(0.057)	***0.201	(0.055)	***0.190	(0.060)
TwoParents	0.085	(0.083)	0.069	(0.081)	**0.168	(0.076)
Separation	0.057	(0.047)	0.050	(0.044)	0.055	(0.045)
PmkImmig	-0.049	(0.059)	-0.050	(0.059)	-0.018	(0.063)
Sibling1	**-0.143	(0.058)	**-0.138	(0.055)	-0.058	(0.059)
Siblings2	*-0.110	(0.057)	*-0.092	(0.054)	-0.021	(0.060)
ComSizeRural	-0.048	(0.043)	-0.063	(0.042)	*-0.074	(0.042)
ComSize2	0.020	(0.038)	0.010	(0.036)	-0.011	(0.037)
SchoolQIndex	-0.002	(0.010)	-0.002	(0.010)	-0.011	(0.011)
Q2FamInc	0.005	(0.056)	0.001	(0.052)	0.035	(0.053)
Q3FamInc	0.041	(0.055)	0.005	(0.054)	0.048	(0.049)
Q4FamInc	0.072	(0.059)	0.056	(0.057)	0.079	(0.056)
Q2Math	***0.152	(0.057)	**0.121	(0.055)	0.068	(0.059)
Q3Math	***0.223	(0.053)	***0.169	(0.052)	*0.096	(0.054)
Q4Math	***0.299	(0.057)	***0.246	(0.056)	***0.196	(0.059)
Q2Reading	0.077	(0.053)	0.057	(0.052)	0.071	(0.051)
Q3Reading	0.060	(0.052)	0.036	(0.049)	0.013	(0.049)
Q4reading	*0.105	(0.058)	0.066	(0.057)	0.052	(0.059)
HealthE&VG	0.048	(0.037)	0.040	(0.037)	0.016	(0.039)
SelfEsteem	0.004	(0.009)	-0.006	(0.008)	-0.016	(0.011)
HyperActInat	-	-	***-0.023	(0.005)	***-0.024	(0.005)
Aggression11	-	-	***-0.022	(0.010)	-0.015	(0.010)
Nurturance	-	-	-	-	**0.008	(0.004)
Constant	0.227	(0.165)	***0.611	(0.160)	***0.645	(0.177)
Observations	1,155		1,145		1,010	
Adj. R^2	0.233		0.283		0.250	

Notes: 1. Bootstrapped standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

For cognitive factors, males who obtained higher scores in their math tests are much more likely to obtain their diploma than those in the bottom of the distribution of math scores. For females math or reading scores barely affect their graduation probability. In preliminary estimations, including the math score in the regression increased substantially the R2. Amazingly, adding reading scores or income quartiles adds almost no additional prediction power to the model for

males. This of course does not mean reading is not important. Math skills feed off reading skills and vice-versa, but knowledge of reading scores is not particularly helpful for prediction once math achievement is known. More information about the link between these measures in Table A5 which presents the matrix of quartiles of reading (from mean in cycles 2 and 3) by the quartiles of math scores (from mean in cycles 1 to 4); and in Table A6 which presents the educational status of the youth by quartiles of math and reading scores of the youth. The only non-cognitive measure used, self-esteem is not significant except in specification (1) for females.

Being healthy (excellent and very good reported health) is a significant and important predictor of graduation for females (.08 impacts in all specification). For males, this auto-declared characteristic is not significant. The two measures of behavioural problems (hyperactivity and aggression) influence negatively and significantly the probability of graduation for both gender, but their impact is small at the margin. However, pathological cases have a much lower probability of graduating. For men, a hyperactive child with the highest score (11) compared to a child with a score of 0, the lowest score, would have a graduating probability that is .27 smaller in specification (3). Finally, the parental nurturance scale is never significant except in specification (3) for males with a marginal impact.

Probability of PSE participation (conditional on high school diploma)

We pursue with the analysis of the probability of PSE attendance for youth who have a HS degree. As explained earlier, we aggregate all types of PSE in order to include the sample from the province of Québec in the regressions. A very small number of individuals declare a PSE with no HS diploma. Therefore we choose to ignore them. Most factors that influence graduation probabilities affect PSE probabilities with the same sign. The PMK education effects are generally larger for females particularly in specification (3) which is particularly revealing as we will see for males. A PMK with the lowest level of education will reduce a young female's probability of PSE by -.37 compared with a PMK with a high school degree, this value is only -.15 for males. It is interesting to note that this coefficient for males drops by -.075 and becomes not significant when nurturance is added in the regression. Therefore, mothers with a very low level of education also display a low level of nurturance for males. The PMK with a university diploma effect is .18 higher for females, with a college degree the difference is .14. Finally, none of the PMK education effects are significant for males, a surprising result. Being raised in a two-parent household relative to a single mother household, as for HS graduation, is very important for both genders coefficients ranging form .08 to .16. Experiencing a separation has no impact on PSE. The same is true for living with an immigrant PMK, the number of brothers and sisters, and the quality of school. There is a sizable negative effect of living in a rural community for men effects ranging from -.09 to -.12. Income effects are very small and far from significant for both males and females in specification (3). Achievement in mathematics has very strong effects for males, effects reaching .14 and .22 for Q3 and Q4 respectively while reading has very small effects that are not significant.

For females, only the Q4 math quartile has a significant effect, relatively large but much smaller than males at around .08. Interestingly the same is true for reading where the effect is approximately the same as in mathematics for Q4. Excellent or very good health also makes an important difference as for graduation from HS. The presence of nurturing reduces the size of the health effect for males while making it not significant expressing a strong link between

nurturance and health which is not surprising. Self-esteem is never significant for both genders but hyperactivity is significant for males and aggression significant for females, with pathological cases seriously reducing the chances of a PSE education even when graduating from HS.

Table 7: OLS estimation of post-secondary education participation if high-school diploma, 18 to 21 year olds women, NLSCY, Canada 2004-2005

Variables	Specifications					
	(1)		(2)		(3)	
Age19	**0.096	(0.043)	0.041	(0.041)	0.033	(0.041)
Age 20	***0.162	(0.043)	**0.098	(0.045)	*0.087	(0.046)
Age 21	***0.202	(0.040)	***0.140	(0.039)	***0.117	(0.041)
Que	**0.093	(0.038)	**0.083	(0.042)	*0.080	(0.043)
Ont	0.028	(0.038)	0.018	(0.039)	0.016	(0.040)
Prai	*-0.080	(0.047)	-0.069	(0.046)	-0.068	(0.047)
Alb	-0.042	(0.046)	-0.027	(0.047)	-0.028	(0.051)
BC	-0.007	(0.051)	-0.005	(0.049)	-0.008	(0.051)
PmkEduPrim	-0.196	(0.167)	-0.270	(0.182)	*-0.326	(0.188)
PmkEduSec	0.006	(0.050)	0.033	(0.055)	0.049	(0.057)
PmkEduSomePSE	0.046	(0.040)	0.049	(0.043)	0.054	(0.044)
PmkEduCol	0.069	(0.042)	*0.080	(0.047)	*0.085	(0.049)
PmkEduUni	*0.078	(0.045)	**0.106	(0.048)	**0.119	(0.050)
TwoBioParents	0.056	(0.053)	0.052	(0.055)	0.076	(0.057)
TwoParents	**0.115	(0.057)	**0.144	(0.058)	**0.156	(0.061)
Separation	-0.052	(0.043)	-0.058	(0.042)	-0.056	(0.045)
PmkImmig	-0.018	(0.042)	-0.006	(0.039)	-0.007	(0.040)
Sibling1	-0.043	(0.040)	-0.009	(0.038)	-0.009	(0.039)
Siblings2	-0.0044	(0.039)	0.017	(0.039)	0.022	(0.040)
ComSizeRural	-0.015	(0.034)	-0.019	(0.035)	-0.015	(0.037)
ComSize2	-0.028	(0.031)	-0.033	(0.032)	-0.021	(0.035)
SchoolQIndex	-0.004	(0.009)	-0.006	(0.010)	-0.004	(0.010)
Q2FamInc	0.039	(0.044)	0.025	(0.044)	0.017	(0.045)
Q3FamInc	0.041	(0.045)	0.028	(0.046)	0.026	(0.046)
Q4FamInc	0.067	(0.046)	0.039	(0.046)	0.023	(0.047)
Q2Math	0.003	(0.045)	0.010	(0.045)	0.023	(0.046)
Q3Math	-0.000	(0.043)	-0.020	(0.045)	-0.024	(0.046)
Q4Math	*0.079	(0.045)	*0.086	(0.045)	*0.083	(0.048)
Q2Reading	**0.100	(0.047)	0.069	(0.049)	0.049	(0.051)
Q3Reading	*0.086	(0.045)	0.056	(0.046)	0.043	(0.046)
Q4reading	***0.130	(0.046)	**0.100	(0.047)	*0.089	(0.048)
HealthE&VG	**0.062	(0.028)	**0.063	(0.031)	**0.066	(0.031)
SelfEsteem	0.003	(0.007)	-0.006	(0.007)	-0.003	(0.007)
HyperActInat	-	-	*-0.008	(0.005)	-0.007	(0.005)
Aggression15	-	-	*-0.028	(0.016)	*-0.032	(0.017)
Nurturance	-	-	-	-	-0.001	(0.003)
Constant	***0.468	(0.118)	***0.690	(0.127)	***0.669	(0.133)
Observations	1,120		1,026		976	
Adj. R^2	0.135		0.131		0.128	

Notes: 1. Bootstrapped standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: OLS estimation of post-secondary education participation if high-school diploma, 18 to 21 year olds men, NLSCY, Canada 2004-2005

Variables	Specifications					
	(1)		(2)		(3)	
Age19	0.063	(0.055)	0.073	(0.054)	0.070	(0.059)
Age 20	0.046	(0.059)	0.046	(0.060)	0.037	(0.068)
Age 21	***0.174	(0.057)	***0.170	(0.058)	**0.148	(0.064)
Que	0.076	(0.063)	0.093	(0.065)	0.088	(0.069)
Ont	0.024	(0.060)	0.029	(0.060)	0.008	(0.064)
Prai	-0.092	(0.072)	-0.091	(0.072)	-0.080	(0.077)
Alb	-0.105	(0.069)	-0.099	(0.068)	-0.093	(0.072)
BC	-0.058	(0.087)	-0.059	(0.085)	-0.078	(0.089)
PmkEduPrim	*-0.272	(0.155)	*-0.267	(0.157)	*-0.305	(0.165)
PmkEduSec	***-0.224	(0.081)	***-0.232	(0.082)	**-0.208	(0.086)
PmkEduSomePSE	-0.066	(0.056)	-0.074	(0.054)	-0.089	(0.062)
PmkEduCol	-0.011	(0.055)	-0.019	(0.053)	-0.022	(0.060)
PmkEduUni	-0.009	(0.059)	-0.010	(0.060)	-0.004	(0.065)
TwoBioParents	0.079	(0.067)	0.061	(0.068)	-0.010	(0.067)
TwoParents	0.049	(0.100)	0.031	(0.100)	-0.058	(0.107)
Separation	-0.042	(0.057)	-0.049	(0.056)	-0.036	(0.062)
PmkImmig	0.029	(0.062)	0.038	(0.064)	0.020	(0.079)
Sibling1	0.002	(0.056)	0.006	(0.058)	0.011	(0.071)
Siblings2	-0.062	(0.066)	-0.061	(0.067)	-0.059	(0.076)
ComSizeRural	**-0.122	(0.048)	**-0.109	(0.048)	*-0.086	(0.050)
ComSize2	-0.054	(0.046)	-0.059	(0.046)	-0.049	(0.048)
SchoolQIndex	0.006	(0.012)	0.007	(0.012)	0.006	(0.014)
Q2FamInc	-0.039	(0.062)	-0.036	(0.063)	-0.062	(0.068)
Q3FamInc	-0.009	(0.059)	-0.002	(0.056)	-0.008	(0.061)
Q4FamInc	-0.002	(0.061)	0.002	(0.060)	-0.016	(0.065)
Q2Math	0.081	(0.072)	0.083	(0.070)	0.089	(0.077)
Q3Math	**0.139	(0.071)	*0.126	(0.070)	*0.141	(0.076)
Q4Math	***0.227	(0.072)	***0.217	(0.071)	***0.221	(0.079)
Q2Reading	0.018	(0.057)	0.000	(0.056)	0.027	(0.065)
Q3Reading	-0.070	(0.061)	-0.097	(0.060)	-0.071	(0.064)
Q4reading	-0.002	(0.061)	-0.025	(0.060)	-0.002	(0.065)
HealthE&VG	0.070	(0.047)	0.070	(0.046)	0.045	(0.047)
SelfEsteem	0.015	(0.010)	0.009	(0.010)	0.001	(0.012)
HyperActInat	-	-	-0.009	(0.007)	*-0.014	(0.006)
Aggression11	-	-	-0.017	(0.013)	-0.012	(0.014)
Nurturance	-	-	-	-	0.005	(0.005)
Constant	*0.364	(0.189)	***0.527	(0.200)	***0.619	(0.237)
Observations	852		846		768	
Adj. R^2	0.143		0.157		0.144	

Notes: 1. Bootstrapped standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Probabilities for Québec's youth

If we restrict the estimations to the sample of Québec's all youth (see Tables A4-A5), and with a dummy variable for gender, the adjusted R2 are higher and the significant variables are almost all the same as in the preceding estimations. Here are some differences. Evidently by construction,

we obtain a significant gender effect on the probabilities. Having a PMK with mid-education levels affect positively and significantly the probabilities of PSE participation. A university level educated parent do not affects the conditional probability of access to PSE. Family structure does not seem to matter, except in one specification where living in an intact two-parent family affect negatively and significantly the conditional probability of access to PSE (a dubious result). Living in mid-size community influences positively the probability of PSE participation, which may reflect that the colleges, outside Montréal are located in every region. Almost all math and reading quartiles of scores have large positive and significant influence on the probabilities of graduation and access to PSE. Self-esteem and health affect the probabilities of graduation and access to PSE in very few specifications. Hyperactivity is the only behaviour affecting moderately, negatively and significantly the same probabilities.

A familiar theme emerges from these two sets of regressions and will be true of the next set of analyses. While income is not found to be a major player, characteristics associated to low income households play a crucial role for educational attainment. Furthermore, these characteristics are true of the household several years before educational decisions are taken. Single motherhood, low level of parental education, and living in rural areas are characteristics of low income households. Also, low level of achievement in mathematics and reading as well as pathological behaviours are more prevalent in low income households. The results suggest that raising income within the margins that are available to governments should have very small effects on educational attainment. We will return to this point in the section on policy recommendations.

5.2 Multinomial estimation results of three outcomes for youth

We now exclude Québec's respondents and discuss the estimates from a multinomial Logit estimation for high school graduates of the probability that they choose not to pursue higher education, choose college, or university. Table 9 presents marginal effects for the same variables used in the preceding analyses, and for both gender. Tables A7 presents the coefficients and the p-value of the bootstrap standard errors.

We start with the young men (right panel of Table 9). For the remaining sections, we will present results only with regressors from specification (3) in the linear probabilities models of the preceding sections. First, PMK education plays a crucial role for PSE attendance. Very low education levels severely reduce the probability of attaining a PSE level. The two lowest education categories increase the probability of not pursuing a higher level (as compared to living with a PMK with a HS diploma) by .367 for a PMK with only a primary level and .316 for a PMK with some secondary. The decrease in probability for PSE levels is split almost exactly in 2 for a PMK with only a primary level, -.186 College (C) and -.181 University (U) and slightly higher for U, -.194, than for C, -.123, for a PMK with some secondary. For young males with a PMK with a higher than secondary level, only those with a university level clearly stand out as this increases the probability of attending university by .248, while reducing the probability of attending college by -.187 and stopping at the HS level by -.061.

The estimates show no impact of family structure on PSE choices for young men except for the variable separation which seriously diminishes the impact on U by -.181, while increasing the probability of stopping at the HS level (SHS) by .073 and C by .108. This result is similar to the

result in Coelli (2005) where a period of unemployment by a parent reduces the probability of PSE attendance. The separation can be interpreted a transition period of low income a few years before PSE choices must be made. The immigration status of the PMK has no effect and neither does the number of siblings. Living in a rural area alters choices considerably as compared to living in a city, reducing the probability of attending U by -.261, while increasing the probabilities of C by .160 and of SHS by .102. Family income is found to have no effect. Proficiency in mathematics is a key variable for PSE choices. Simply the effect of being in the second quartile rather than the first is huge for young men, as the probability of attending U is .367 greater. The effect on the probability of attending U of moving to the next two quartiles is considerably smaller than moving from quartile 1 to 2 but remains impressive, as the effect of moving from Q2 toQ3 is .076 and from Q3 to Q4, .125. The reading variables are not significant except for the probability of attending C (compensated by an increase of .155 for U). Finally, for health, behavioural, non-cognitive skills and family related variables, only hyperactivity matters as an additional unit in the score will reduce the probability of U attendance by -0.024 and increasing the probability of SHS by almost the same amount. Since the marginal effects are computed at the mean values of the regressors with the mean hyperactivity score at 4.7, and that this score goes to 16, extreme cases of hyperactivity can have vary damaging prospects on a HS graduate's chances of pursuing higher education.

We pursue our analysis with young women (left panel of Table 9). First, PMK university education is the only PMK education variable that plays a crucial role for PSE attendance. This increases the probability of attending university by .306, while reducing the probability of attending college by -.227 and stopping at the HS level by -.080. This is very similar to males.

The estimates show a very strong impact of family structure on PSE choices for young women including the variable separation which diminishes the probability of attending U by almost the same amount as for young men. However, being in a two parent family has a very large positive impact 0.306 on the probability of attending U compared to young females from single-mother homes. Being in a single mother increases the probability of SHS by .140 compared to being in a two parent biological family. The immigration status of the PMK has a vary large effect on the probability of choosing U, 0.210, while reducing by about the same amount the probability of attending C. Living in a rural area alters choices considerably as compared to living in a city, reducing the probability of attending U by -.168, while increasing the probabilities of C by .124 and of SHS by .044. Family income is found to have no effect, like the case of young men. Proficiency in mathematics is again a key variable for PSE choices. The effect of being in the second quartile rather than the first is large but considerably smaller than for young men, as the probability of attending U is .159 greater. The effect on the probability of attending U of moving to the next two quartiles is considerably smaller than moving from quartile 1 to 2, from 2 to 3 the effect is small, .035, but from Q3 to Q4, the increase is very large, .212. The only significant reading variable is Q4, increasing the probability of U attendance by .206, a substantial effect. This increase is done almost totally at the expense of the probability of attending C. Finally, for health, behavioural, non-cognitive skills and family related variables, three variables stand out. First, very good or excellent health increases relative to less than very good health the probability of attending U by .240, while decreasing the probability of SHS by -0.061. Second, the measure of aggression for 11 to 15-year-olds has a very large negative effect on the probability of attending U, -0.081 while hyperactivity increases the probability of SHS, although the effect is

small at 0.006, it can be large if one compares a pathologically hyperactive and an average young woman in the sample. Finally, aggression is a key player for U attendance.

Table 9: Marginal effects of multinomial estimation of PSE participation, high school graduates, 18 to 21 year-olds by gender, NLSCY, Canada except Québec, 2004-2005

Variables	Women			Men		
	Outcomes			Outcomes		
	No PSE	College	University	No PSE	College	University
Age19	-0.022	0.009	0.013	-0.075	0.208***	-0.132
Age 20	-0.081***	0.121	-0.041	-0.086	0.137	-0.051
Age 21	-0.081***	0.100	-0.019	-0.179***	0.195**	-0.015
Ont	0.013	0.248***	-0.261***	0.001	0.033	-0.033
Prai	0.075	0.158**	-0.233***	0.096	-0.041	-0.055
Alb	0.029	0.485***	-0.514***	0.131***	0.070	-0.201***
BC	0.0171	0.276***	-0.293***	0.098***	0.180*	-0.278***
PmkEduPrim	0.184	-0.172	-0.012	0.367*	-0.186	-0.181*
PmkEduSec	0.017	0.070	-0.087	0.316***	-0.123*	-0.194***
PmkEduSomePSE	-0.009	0.001	0.008	0.141*	-0.067	-0.075
PmkEduCol	-0.030	-0.006	0.036	0.058	-0.093	0.0355
PmkEduUni	-0.080***	-0.227***	0.306***	-0.061	-0.187	0.248**
TwoBioParents	-0.140**	-0.163*	0.303***	-0.009	0.065	-0.056
TwoParents	-0.095***	-0.240***	0.334***	0.051	-0.063	0.012
Separation	0.092*	0.084	-0.175*	0.073*	0.108	-0.181***
PmkImmig	-0.018	-0.192***	0.210***	-0.059	0.044	0.015
Sibling1	0.018	0.047	-0.065	0.034	-0.077	0.043
Siblings2	-0.011	-0.013	0.024	0.114	-0.016	-0.098
ComSizeRural	0.044	0.124	-0.168**	0.102***	0.160**	-0.261***
ComSize2	0.047	0.054	-0.100	0.053	-0.059	0.006
SchoolQIndex	-0.000	-0.038**	0.038*	-0.004	0.049	-0.044*
Q2FamInc	-0.011	0.105	-0.094	0.044	0.037	-0.081
Q3FamInc	0.000	0.128	-0.128	-0.018	0.038	-0.020
Q4FamInc	0.001	0.052	-0.052	-0.029	0.093	-0.064
Q2Math	-0.037	-0.122*	0.159**	-0.149***	-0.218**	0.367***
Q3Math	0.021	-0.215***	0.194**	-0.173***	-0.271***	0.443***
Q4Math	-0.153***	-0.254***	0.406***	-0.254***	-0.314***	0.568***
Q2Reading	-0.028	-0.137**	0.165**	-0.028	0.003	0.024
Q3Reading	-0.030	-0.100	0.130	0.063	-0.218***	0.155
Q4reading	-0.044	-0.162**	0.206**	0.003	-0.091	0.088
HealthE&VG	-0.061**	-0.179***	0.240***	-0.062	-0.037	0.099
SelfEsteem	-0.003	0.008	-0.005	0.006	0.003	-0.009
HyperActInat	0.006*	0.008	-0.015	0.025***	-0.001	-0.024**
Aggression15/11	0.026**	0.055**	-0.081***	0.011	-0.023	0.0126
Nurturance	-0.001	-0.007	0.008	-0.009*	0.002	0.007
Observations	809			664		

Notes: Statistical significance from p-value of non bootstrapped standard errors: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To resume, for both young females and males, PMK education, experiencing a separation, the size of the residential community, math proficiency and hyperactivity matter with nuances found

in the text. In contrast, family structure, immigration status, reading scores, health and aggressive behaviour matter for females. Therefore females are sensitive to a much larger number of factors than males when deciding to pursue higher education. The multinomial Logit shows that certain variables that are not significant for PSE in the linear probability model are significant in the Multinomial Logit. For example, separation has no effect in the linear probability model, but has an impact on the type of PSE. Or hyperactivity for females, which has no impact in the linear probability model for females, but increases the probability of SHS.

The theme of "no income effect but characteristics linked to low-income households being determinant" is true of this section as well but another dimension is added from the multinomial approach. These characteristics are negatively related to the probability of attending both C and U but the negative effects are particularly strong for the probability of attending U. This accentuates income inequalities even more because U graduates make higher earnings than C graduates.

Four outcomes for all youth

Before presenting the results from a sequential model, we propose a prediction exercise that may be useful for policy analysis. Given that all the variables we use are from the time the youth is 15 or under, we can predict differences between children on the basis of a simple regression model. For ages 18 to 21, we suppose the youth can be in 4 different states, HS drop-out, HS continuer, SHS, or PSE participant. We then estimate a multinomial Logit model, assuming no HS degree is the base category, and compute marginal effects, these can be interpreted as the predicted differences based on known characteristics of a child when he or she is 15 or under, between two children who have the same observed characteristics except one. For example, the predicted differences in the probability for the 4 states between a child who lives in a rural community and a child who lives a community with a population over 100,000 inhabitants are given by the marginal effects of rural residence. This model is estimated strictly for prediction purposes as a structural choice model supposes that SHS and "HS drop-out" are alternatives, but "HS drop-out" cannot be an alternative if an individual has his HS degree.

Table 10 presents the multinomial Logit results with the 4 possible states for both men and women. Tables A8 and A9 present the coefficients and the p-value of the bootstrap standard errors. We notice that the effects on PSE attendance are larger than for the OLS conditional model because they capture the effect on the choice of graduating and then the effect on PSE choice also all variables affect both choices in the same direction. For example, in the conditional OLS model for females, the lowest PMK education level has a marginal effect on PSE of -.326 in specification (3) for women when conditioned on HS graduation, here we find an effect of -.375. The marginal effect for the highest level of education for the PMK for females is .119 for the conditional case, it is .145. For health, the conditional effect is .065, here it is .111. This regression reinforces the conclusions in the preceding section and makes even more evident the negative impact of certain variables on the probability of PSE participation. The Sequential Logit will add more power to these conclusions with a model that is consistent with utility maximization (in a static setting).

Table 10: Marginal effects of multinomial estimation of four schooling outcomes, youth 18 to 21 year olds by gender, NLSCY, Canada, 2004-2005

Variables	Women				Men			
	Outcomes				Outcomes			
	(1) HS Drop-out	(2) HS continuer	(3) HS graduate non-PSE	(4) PSE participant	(1) HS Drop-out	(2) HS continuer	(3) HS graduate non-PSE	(4) PSE participant
Age19	0.052*	-0.044***	-0.016	0.008	-0.014	-0.054***	-0.038	0.106**
Age 20	-0.018	-0.056***	-0.047**	0.120***	-0.028**	-0.070***	-0.012	0.110**
Age 21	0.002	-0.068***	-0.068***	0.135***	-0.024*	-0.109***	-0.109***	0.242***
Que	0.036	-0.023**	-0.073***	0.061	0.132**	0.046**	-0.131**	-0.046
Ont	0.031	0.0045	-0.020	-0.015	0.018	0.041	-0.017	-0.042
Prai	-0.005	-0.0089	0.037	-0.023	0.053	0.000	0.051	-0.104
Alb	0.034	-0.016	0.004	-0.022	0.026	0.019	0.099	-0.144*
BC	0.008	-0.007	0.012	-0.013	0.034	-0.033**	0.083	-0.084
PmkEduPrim	0.184	0.019	0.172	-0.375***	0.040	-0.027*	0.169	-0.182
PmkEduSec	-0.018	-0.009	-0.032	0.059	0.019	-0.026**	0.196**	-0.189**
PmkEduSomePSE	-0.015	-0.010	-0.037	0.062*	-0.008	-0.025*	0.083	-0.051
PmkEduCol	-0.026**	-0.029***	-0.049*	0.104***	-0.027*	0.003	0.012	0.012
PmkEduUni	-0.037***	-0.037***	-0.072***	0.145***	-0.033**	0.014	-0.008	0.028
TwoBioParents	-0.038	-0.014	-0.029	0.081*	-0.106*	-0.068*	0.044	0.130
TwoParents	-0.018	-0.019*	-0.068***	0.104***	-0.018	-0.040***	0.068	-0.009
Separation	-0.002	-0.011	0.033	-0.020	-0.016	-0.004	0.049	-0.029
PmkImmig	-0.018	-0.009	0.002	0.025	0.063	-0.007	0.001	-0.058
Sibling1	-0.039**	-0.013	0.025	0.028	0.071**	-0.024	-0.023	-0.024
Siblings2	-0.024	-0.009	0.011	0.022	0.051	-0.031	0.048	-0.068
ComSizeRural	-0.010	-0.007	0.005	0.011	0.044	0.019	0.076	-0.139**
ComSize2	-0.009	0.021	0.012	-0.024	-0.015	0.020	0.051	-0.056
SchoolQIndex	-0.003	-0.005	0.003	0.005	0.003	0.003	-0.009	0.003
Q2FamInc	-0.019	-0.010	-0.007	0.036	-0.019	0.037	0.061	-0.080
Q3FamInc	-0.047***	0.003	-0.018	0.062*	-0.019	0.001	0.013	0.005
Q4FamInc	-0.010	-0.010	-0.017	0.046	-0.021	-0.002*	0.008	0.016
Q2Math	-0.013	0.004	-0.009	0.017	-0.014	-0.011	-0.039	0.064
Q3Math	-0.026**	0.014	0.023	-0.011	-0.012	-0.026*	-0.077	0.115**
Q4Math	-0.031**	0.016	-0.070**	0.084*	-0.057***	-0.045***	-0.138**	0.241***
Q2Reading	-0.015	-0.014	-0.034	0.064**	-0.014	-0.017	-0.029	0.060
Q3Reading	-0.011	-0.007	-0.028	0.046	0.019	-0.011	0.056	-0.064
Q4reading	0.000	-0.019	-0.067**	0.085**	-0.001	-0.030*	-0.008	0.040
HealthE&VG	-0.036**	-0.025*	-0.050*	0.111***	-0.014	-0.011	-0.031	0.056
SelfEsteem	0.000	-0.002	0.005	-0.003	0.002	0.001	0.002	-0.006
HyperActInat	0.000	0.005***	0.005*	-0.010***	0.006***	0.004***	0.010	-0.020***
Aggression15/11	0.013***	0.006	0.014*	-0.033***	0.004	0.008**	0.006	-0.019
Nurturance	-0.002*	0.003**	-0.000	-0.001	-0.003**	-0.000	-0.004	0.007
Observations	1,168				1,010			

Notes: 1. Statistical significance from p-value of non bootstrapped standard errors for marginal effects: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. HS: High-school; PSE: postsecondary education.

6. Self-selection and sequential Logit results

Figures 1 and 2 present a general framework of schooling transitions showing that they are conditional on school "success", that is grade/level transition. The exit means labour market activities or other activities than schooling. Nevertheless, we do observe that a small number of youth are enrolled in college (or university) without having a high school diploma. Moreover, the transition between college and university is a possibility. Especially in Québec, where the college level is in fact divided in two "path": "technical/professional" college" (3 years) and "general" college (2 years) which is prerequisite for university access.

Does Self-Selection matter?

As explained in the introduction, conditioning on graduation from high school to estimate PSE attendance choice models, may produce biased estimates of the PSE model. Because high unobserved ability individuals are over represented for certain groups conditional on high school graduation (e.g., high-income families, high math score) we can expect a negative correlation, condition on high school graduation, of human capital variables and unobserved variables in the PSE model. We estimate jointly a graduation from high school Logit and a PSE attendance Logit; assuming unmeasured characteristics are in part child-specific and constant across the high school graduation and PSE attendance choices. For estimation purposes, we assume that the child specific unobserved attribute has a finite mixture distribution[30] with two types of individuals, which can be interpreted as high- and low-ability individuals. Because we add a constant in the model, only one point of support is estimated with its associated weight.

Because of the two year difference between waves, it is difficult to know whether children who are in the twelfth grade in wave 5 of the NLSCY attended a PSE institution in the year following graduation and then dropped out. Therefore, we perform our analysis with children who are in the eleventh grade in cycle 5. We also excluded children from Ontario. We find the graduation rate to be unusually low for this group. This could be caused by the double cohort phenomenon. Finally, we exclude children from Québec with a very high PSE attendance diminishing the selection bias issue. We present in Table 11 the results for the math score quartile effects and income quartile effects for a Logit conditional on graduation and the results from our joint estimation of a Logit for graduating high school and a Logit for PSE attendance with a finite mixture distribution for the child specific random effect. Other explanatory variables are: gender, good health, living in rural area, hyperactivity score, regional dummies, PMK education and age.

Richer specifications are much more tedious and take much more time to estimate, given time constraints, we performed the analysis with a smaller set of regressors. However, we believe the conclusions about selection bias would not change with more regressors.

---

[30] In statistics, a **probability mixture model** is a probability distribution that is a convex combination of other probability distributions. Suppose that the discrete random variable X is a mixture of n component discrete random variables Y_i. Then, the probability mass function of X, $f_X(x)$, is a weighted sum of its component distributions:

$$f_X(x) = \sum_{i=1}^{n} a_i f_{Y_i}(x)$$

For some mixture proportions $0 < a_i < 1$ where $a_1 + \cdots + a_n = 1$. The definition is the same for continuous random variables, except that the functions f are probability density functions.

Table 11: Finite mixture and standard Logit estimations

Explanatory variables	Finite mixture*	Standard Logit
Q2FamInc C1_4 (Q1)	0.760	0.666
Q3FamInc C1_4	1.000	0.817
Q4FamInc C1_4	0.660	0.510
Q2Math 7_15 (Q1)	0.714	0.630
Q3Math 7_15	1.273	1.106
Q4Math 7_15	1.878	1.668

* For the finite fixture model, the explanatory variables are identical in both the Logit for graduation from high school and PSE attendance.

As expected, there is a slight negative bias with the Logit conditional on high school graduation for math and income effects. Therefore, our results can be considered a lower bound for these particular effects. The support point is estimated with a very low t-statistic, providing evidence that there is no selection bias problem for PSE attendance models conditional on high school graduation. Therefore, the estimation with a simple sequential Logit model with unobserved heterogeneity will provide credible results for policy analyses.

A simple sequential Logit model

Tables 12 and 13 presents of the sequential Logit estimated parameters of schooling transition for the 19- to 21-year-old women and men respectively, in all provinces except Québec. The decision tree estimated is the following: the first branch is graduating HS versus not graduating; the second is conditional on graduation do they chose to stop or move to university or college. Because the sequential Logit is estimated with the same youth for both decisions, we chose not to include the 18 year-olds as too many are in high school, therefore the high school graduation outcome is not observed for several of these. Of course, for several of the 19-, 20- and 21-year-olds, their final decision concerning their final level of education is not taken, implying that we should concentrate on the 21 year-olds, the closest to the final decision. In this case, the sample of size would have been very small. So we decided to include youth older than 18, mainly for reasons of sample size.

The first decision is binary; therefore, it is a Logit. Given this decision, the second involves three possibilities (stop, college, or university), hence, a multinomial Logit (MNL). In tables 10 and 11 the results show in the first two columns the odds ratio estimates and p-values respectively for the Logit of HS graduation. In columns 3 and 5, are displayed the odds ratio for the MNL with no PSE as the baseline category.

The results are very similar of course to the OLS results (in Tables 6 and 8), and the MNL results (in Table 9), but much fewer parameters are significant because of smaller sample size. Therefore, we shall not discuss the results for these Logits. However, we estimate the total effect of the variables on the probability of not graduating, of stopping after HS graduation, of attending college, or of attending university. Call β_{HS} the row vector of coefficients from the first Logit of graduation and β_i, those for the MNL estimation, i=C, or U; let x be the column vector of explanatory variables. Then, the total probabilities of not graduating (NG), of stopping after HS graduation (SHS), of attending college (C), or of attending a university (U) are respectively given by:

$$P_{SHS} = P_{HS}(x)(\frac{1}{1+e^{\beta_C x}+e^{\beta_U x}}),$$

$$P_C = P_{HS}(x)(\frac{e^{\beta_C x}}{1+e^{\beta_C x}+e^{\beta_U x}})$$

$$P_U = P_{HS}(x)(\frac{e^{\beta_U x}}{1+e^{\beta_C x}+e^{\beta_U x}}) \quad (1)$$

$$P_{NG} = 1 - \frac{e^{\beta_{HS} x}}{1+e^{\beta_{HS} x}}$$

Where, P_{HS} is the probability of graduating from HS. To estimate the marginal effects, we start with a baseline case such that the probability of graduating from high school is approximately equal to the percentage of young men or young women depending on the sample and we simply change the value of one regressor and recomputed the probability, the difference between the new probability and the baseline case is the marginal effect; for the marginal effect of continuous regressors we fix the value at the mean of the variable for the sample. We then estimate the effect of a continuous variable by increasing it by one standard deviation, re-computing the baseline probability with this new value and take the difference between this probability and the baseline probability. The results appear in Table14. We compute the effects only for variables that were significant or close to being significant in the sequential Logit. The baseline case is a person from Ontario who is 21 years old with a PMK with a high school degree but no more, with two biological parents, 1 sibling, living in a large city in the second income quartile, second math quartile, third reading quartile, in excellent/very good health with continuous variables evaluated at the mean.

Not reporting being in excellent or very good health has a very large effect on the probability of being a drop-out for young men, increasing this probability by .25 and reducing the probability of choosing to SHS, choosing C, and U, by respectively -.04, -.06 and -.16. Interestingly, for female the main effect is to decrease the probability of attending U by -.16 and increasing the probability of attending C by .14, therefore "poor health" affects the type of PSE but not the probability of attending a PSE level institution.

Having a very poorly educated PMK has dramatic consequences for both young men and women, when compared to a youth with a PMK with a HS degree as highest level of education, as it increases the probability of being a drop-out by .61 for females and .44 for males. The proportion of children in this situation is very low, but large enough to be alarming. Again, for both sexes, a university educated PMK will have an enormous impact compared to a mother with a HS degree. For young women, the main impact is to increase the probability of U attendance by .41 and decreasing the probability of attending C by -.33. For young men, the increase of the probability of U attendance is .40, but the decrease in the other choices are spread out, -.17 for not graduating, -.11 for SHS, and -.12 for C.

In terms of family structure, experiencing a separation is particularly deleterious to a university level education for both sexes, however it has little overall affect on PSE attendance. A single-mother home compared to a two biological-parent home has dramatic effects on schooling choices. For females, we observe a .11 rise in the probability of drooping out and a -.21 decrease in the probability of attending U, for males, the increase in the probability of dropping out is .37, which is very large.

Moving from a Q2 math ranking to a Q4 ranking, have dramatic impacts on the probability of U attendance, increasing it by .28 for young women and .26 for young women. For reading, effects are less substantial, but remain important. Moving from a Q3 ranking to a Q1 ranking increases the probability of C attendance by .10 and decreases the probability of U attendance by .12, therefore little impact on the probability of PSE attendance, for young men the same is somewhat true, but the large decrease in the probability of being a drop-out in favour of attendance in C is not credible and is probably the result of the co-linearity of this variable with others. It is a rare instance of a counterintuitive result for this paper.

Hyperactivity strongly affects the choice of males but not females. An increase of one standard deviation of the hyperactivity score with respect to the mean decreases the probability of attending U by -.12 while increasing the probability of SHS by .8, which is considerable. A one standard deviation in the aggression score around the ages of 8-11 years seriously diminishes the probability of attending university and increases the probability of dropping out or by .05. We do not compute the income effects as they are far from significant.

Table 12: Sequential Logit estimation of schooling transition, 19 to 21 year olds women, all provinces except Québec, 2004-2005

Decision tree	High school graduate		College		University	
Control variables	Odds ratio	P-value	Odds ratio	P-value	Odds ratio	P-value
Age 20	8.94	0.00	2.98	0.01	1.89	0.17
Age 21	1.43	0.53	2.86	0.02	2.07	0.12
Ont	0.35	0.22	2.05	0.14	0.49	0.15
Prai	1.94	0.52	0.93	0.89	0.27	0.02
Alb	0.32	0.24	3.54	0.02	0.16	0.01
BC	0.98	0.99	2.50	0.18	0.51	0.30
PmkEduPrim	0.09	0.01	0.17	0.06	0.29	0.14
PmkEduSec	1.43	0.65	1.03	0.97	0.82	0.77
PmkEduSomePSE	1.96	0.23	0.90	0.82	0.88	0.77
PmkEduCol	2.30	0.18	1.20	0.73	0.98	0.97
PmkEduUni	4.44	0.10	0.84	0.81	4.90	0.02
TwoBioParents	2.42	0.19	1.73	0.35	5.91	0.01
TwoParents	0.71	0.69	2.00	0.32	12.58	0.00
Separation	1.86	0.35	0.66	0.41	0.27	0.04
PmkImmig	1.17	0.77	0.43	0.10	1.51	0.43
Sibling1	7.30	0.00	0.63	0.36	0.53	0.27
Siblings2	4.72	0.01	1.00	1.00	1.12	0.85
ComSizeRural	1.82	0.35	1.15	0.79	0.56	0.25
ComSize2	1.23	0.71	0.72	0.40	0.52	0.12
SchoolQIndex	1.10	0.40	0.77	0.04	1.01	0.92
Q2FamInc	2.09	0.25	1.94	0.14	1.36	0.53
Q3FamInc	1.22	0.00	1.76	0.33	0.91	0.87
Q4FamInc	2.97	0.15	1.15	0.83	1.04	0.95
Q2Math	0.79	0.70	0.97	0.94	2.01	0.13
Q3Math	2.47	0.22	0.35	0.03	1.12	0.82
Q4Math	1.88	0.40	3.44	0.31	2.70	0.01
Q2Reading	0.44	0.29	0.74	0.50	2.11	0.11
Q3Reading	0.69	0.60	1.45	0.43	2.79	0.04
Q4reading	0.63	0.51	0.72	0.54	2.36	0.11
HealthE&VG	1.23	0.68	1.06	0.87	3.23	0.00
SelfEsteem	1.10	0.34	1.10	0.23	1.06	0.52
HyperActInat	0.98	0.72	0.09	0.14	0.89	0.02
Aggression15	0.83	0.09	0.90	0.37	0.61	0.00
Nurturance	1.11	0.04	1.00	0.97	1.03	0.47
Observations	687					
Wald chi2(34)	136.52; Probability > chi2 = 0.0000					

Notes: 1. Robust standard error; .sstatistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1. Weighted estimations and bootstrapped standard errors.

2. Decision tree (0:1 2 3; 1: 2: 3), where: Drop-out of HS = 0; HS graduate but non-PSE participant = 1; College participant = 2; University participant = 3.

Table 13: Sequential Logit estimation of schooling transition, 19-21 year olds men, all provinces except Québec, 2004-2005

Decision tree	High school graduate		College		University	
Control variables	Odds ratio	P-value	Odds ratio	P-value	Odds ratio	P-value
Age 20	2.05	0.31	1.02	0.96	1.61	0.29
Age 21	1.83	0.28	1.79	0.14	2.59	0.04
Ont	0.57	0.45	1.05	0.92	0.71	0.55
Prai	0.32	0.08	0.53	0.28	0.68	0.57
Alb	0.68	0.62	0.65	0.47	0.25	0.03
BC	0.51	0.50	0.74	0.64	0.20	0.02
PmkEduPrim	0.12	0.07	0.28	0.26	0.12	0.13
PmkEduSec	0.19	0.07	0.15	0.00	0.16	0.02
PmkEduSomePSE	0.93	0.93	0.35	0.03	0.36	0.10
PmkEduCol	9.78	0.02	0.57	0.28	1.18	0.78
PmkEduUni	5.08	0.00	0.41	0.22	1.66	0.49
TwoBioParents	5.74	0.02	1.08	0.89	1.11	0.87
TwoParents	1.27	0.80	0.51	0.38	1.22	0.80
Separation	2.56	0.20	1.30	0.52	0.42	0.14
PmkImmig	0.21	0.05	1.38	0.57	1.30	0.64
Sibling1	0.24	0.06	0.89	0.82	1.32	0.63
Siblings2	0.30	0.10	0.68	0.47	0.57	0.35
ComSizeRural	0.58	0.48	0.80	0.60	0.14	0.00
ComSize2	1.62	0.44	0.57	0.13	0.79	0.60
SchoolQIndex	1.03	0.91	1.12	0.31	0.85	0.19
Q2FamInc	2.77	0.13	0.90	0.82	0.55	0.32
Q3FamInc	1.51	0.52	1.55	0.33	0.75	0.63
Q4FamInc	0.40	0.35	1.18	0.76	0.65	0.52
Q2Math	2.32	0.14	1.72	0.19	9.14	0.00
Q3Math	2.32	0.28	1.25	0.61	9.23	0.00
Q4Math	10.98	0.02	1.95	0.23	23.45	0.00
Q2Reading	1.58	0.53	2.00	0.12	1.86	0.32
Q3Reading	0.37	0.13	0.52	0.12	1.77	0.28
Q4reading	0.44	0.37	1.46	0.42	1.89	0.30
HealthE&VG	3.63	0.03	1.07	0.84	1.36	0.48
SelfEsteem	0.90	0.33	0.99	0.93	0.92	0.46
HyperActInat	0.95	0.57	0.92	0.11	0.80	0.00
Aggression11	0.78	0.07	0.98	0.87	1.03	0.78
Nurturance	1.11	0.01	1.04	0.21	1.06	0.16
Observations	561					
Wald chi2(34)	2972.99; Probability > chi2 = 0.0000					

Notes: 1. Robust standard error; .sstatistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1. Weighted estimations and bootstrapped standard errors.

2. Decision tree (0:1 2 3; 1: 2: 3), where: Drop-out of HS = 0; HS graduate but non-PSE participant = 1; College participant = 2; University participant = 3.

Table 14: Sequential Logit base probabilities and simulations

Decision tree	Schooling transition				Schooling transition			
	Dropout of HS	HS grad	College	University	Dropout of HS	HS grad	College	University
	Women				Men			
Estimated base probabilities	10.0	6	53	31	17	20	25	38
Simulations results								
Health_56=0	+2	+2	+14	-18	+25	-4	-6	-16
PmkEduPrimary=1	+61	0	-42	-20	+44	-5	-10	-29
PmkEduUniversity=1	-5	-3	-33	+41	-17	-11	-12	+40
Two bio-parents=0	+11	+5	+5	-21	+37	-8	-11	-18
Separation=1	-4	+5	+14	-15	-11	+7	+19	-16
Q4Math 7-15=1	-6	-5	-18	+28	-13	-7	-6	+26
Q1Reading=1	-3	+4	+10	-12	-10	+1	+25	-16
Hyperactivity+1SD	+1	+2	+1	-3	+2	+8	+2	-12
Agression15/11+1SD	+3	+2	+7	-11	+6	-2	-3	-2
Nurturance+1SD	-1	0	0	+1	-1	0	0	+1

Notes: HS is high school and grad is a high school graduate. The baseline case is a person from Ontario who is 21 years old with a PMK with a high school degree but no more, with two biological parents, 1 sibling, living in a large city in the second income quartile, second math quartile, third reading quartile, in excellent/very good health with continuous variables evaluated at the mean.

7. Policy implications of results

Methodologically, it is, we believe, important to perform the analysis by gender as the results differ considerably. For example, levels of aggression have no impact for young men, while the level of aggression as measured with a scale built for children between 11 and 15 years has an impact on young women. However, a large number of variables have impacts that are similar for both gender, but the differences in the impacts are large enough to warrant separate regressions for both groups. This necessary leads to thinking about gender gaps: the explanation as why young men are losing ground proves elusive. The case of Québec demonstrates that this is a complex problem. Although young men from Québec perform just as well as those from the rest of Canada in reading and math, their high school graduation rate is much lower (but college participation is high if high school graduates), demonstrating that factors other than cognitive qualifications are driving this result. Given high school graduation, more women choose university than men. Math scores are hugely important for university attendance for males who are in the first quartile which is less the case for women while reading scores are important for university attendance for females while males are insensitive to reading scores. This could be related to the type of academic programs that interest young males and females. The YITS is a better data set to address the issue of academic program preference. The absence of reading effects for young males is an intriguing result. It must be noted that our measure of reading covers only 2 cycles, while our math measure covers 4 cycles and therefore could be more precise.

Despite the fact that income, as measured as the mean income ($2002) of the family during cycles 1 to 4, does not seem to be a key player for PSE attendance or high school graduation, the sign of its effect is generally positive and non-linear, increases for children in very low income will have a larger effect that those with higher levels. More importantly, several variables that are characteristics of low-income families play a key role for schooling attainment. For example, being from a single-parent home with a poorly educated PMK and with less than (perceived) excellent/very good health or with high levels of hyperactivity for males or high levels of aggression for young teenage females will almost negate any chance of attaining the level of PSE. These results are now becoming familiar through this literature, i.e. despite the fact that being in a low-income family reduces the chances of attaining PSE, increasing the income of these families should have a minor effect on the probability of attaining a level of PSE.

We find that developing cognitive abilities, good health and controlling hyperactivity and aggression are more important than income for PSE. It is also known that ability (cognitive and non cognitive) gaps open up early in life, that to chronic health problems early in life is conducive to a poor health status and that aggressive and hyperactive behaviours are identifiable at a very early stage in child development. Also, some periods are critical for development while other periods are sensitive, to use Cunha and Heckman's (2007a, 2007b) terminology.

Therefore, policy must be designed to address the problems that are associated with poorly educated parents or single parent homes and the types of interventions that seem to work best are those which are aimed at improving the lot of children as early as possible in their lives. Clearly, children benefiting the most from these interventions are from low-income families as they have the most to gain. On average, children from middle- and high-income families with two-parents are doing very well. Also, we have seen that hyperactivity and aggression have negative impacts

on schooling attainment and work by Tremblay et al. has shown that interventions that are made very early can considerably dampen any pathological trajectory of aggression.

Early childhood interventions in the life cycle of disadvantaged children offer a higher return than later interventions such as better schools, because they offer benefits very early. For example, school preparedness reduced the need for remedial programs which are very costly. Chronic health problems developed early in life are very costly. High levels of aggression developed early in life can lead to criminal behaviour in the future. Too many children start grade school with a very small chance of attending PSE.

This does not mean that we should not help low income students with funding for higher education. It simply means that within a portfolio of interventions geared to helping children attend PSE, proportionally, more resources should be devoted to early childhood interventions targeted towards children from low income households.

A surprising result is that better schools do not seem to have an impact on PSE attendance. However, in recent work Lefebvre, Merrigan and Verstrate (2008a) show that high quality schools have a positive impact on math scores. Therefore, although our regressions show no direct impact on PSE, it is possible that there is an indirect effect on PSE channelled by higher achievement in mathematics. In related work, Lefebvre, Merrigan and Verstrate (2008b), find that attending a private high-school in Québec, which are heavily subsidized by the provincial government and creating a competitive environment for the schooling system, has a strong positive effect on math achievement. The study carefully addresses the issue of simultaneity between private school attendance and achievement in mathematics. Creating a more competitive environment within the schooling sector is achievable by public policy.

Given these facts, it is more difficult for the federal government to directly intervene in this area. Its main tools fiscal policies and transfers cannot change the picture of PSE in any important way in Canada. However, there are other ways the government can act. Although some experimental studies have shown that early interventions can change the future lives of young children, the evidence is mostly American. The federal government should be funding more experimental studies that seek to improve the cognitive and behavioural development of young children. There is no quick fix fiscal policy that can substantially increase the chances of children with the characteristics we have identified as very detrimental to a child's chance of graduating from high-school or receiving a PSE.

8. Conclusion

We start pointing out the limits of our study. First, several of the children in the cohort we use for the empirical analysis have not completed their education (10.2% of the weighted sample of 1,509,944 respondents). Second, the level of attrition in cycle 6 compared to cycle 5 is very high. Fortunately, in cycle 7, we will observe approximately 6,300 respondents 18-23 year-olds. Data for the YITS show that schooling status for a cohort of 22-24 is more representative of the final level of educational attainment. Therefore, work with cycle 7 should provide more "final answers" to the questions surrounding PSE. The non-response rate for math and reading scores as well as non-cognitive skills is high leading to possible selection bias in the results.

In this paper, we tried to find the major determinants of PSE for young Canadian males and females, using the only Canadian data set, the NLSCY, containing a lengthy history of childhood experiences as well as measures of different aspects of their psychological and socio-emotional development. We were able to determine that mathematical skills observed from the age of 8 up to the age of 15 were key players for PSE choices. We found that reading skills did not have an impact of the same size particularly for young men. We also found that a very high level of hyperactivity at the age of 11 is a serious handicap for PSE attendance. The same is true for the aggressive behaviour of females as measured when they are early teens (11 to 15). Non-cognitive variables found to be important as in the work of Cunha and Heckman were not found to be statistically significant in our study (partial response might have played a role here). This could be explained by the methodology used by Cunha and Heckman, as they used factor analysis to merge the two non-cognitive measures into a single latent factor. Future research should try the Cunha and Heckman method for the same analysis. Variables pertaining to the quality of neighbourhoods, parental practices, and family functioning, and the quality of schools attended by children were found either not statistically significant or significant but with a marginal impact.

Most children with pathological levels of hyperactivity or aggressiveness are from low-income families and develop these behaviours in early childhood. The same is true of less than very good health and low cognitive skills. Hence, we return to the opinion that the most efficient policies are in-kind, such as direct interventions by child development professionals, and are those that concentrate on young children in disadvantaged families. Regarding this conclusion, economists and the rest of social scientists and health care specialists in the area are at odds, as the latter believe a "universal" approach will yield better results than targeting.

Figure 1: A simplified educational system (all provinces except Québec)

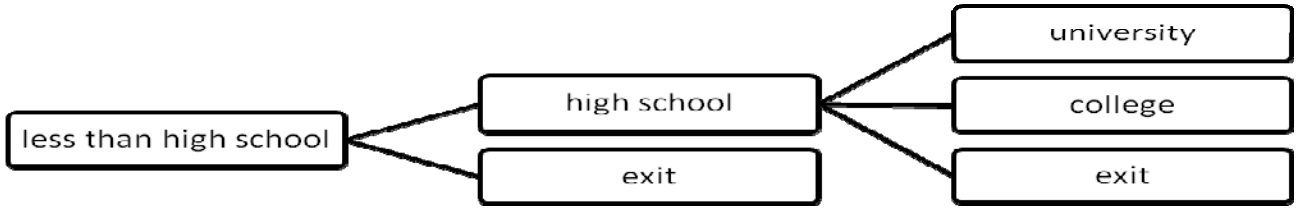

Figure 2: A simplified educational system (Québec only)

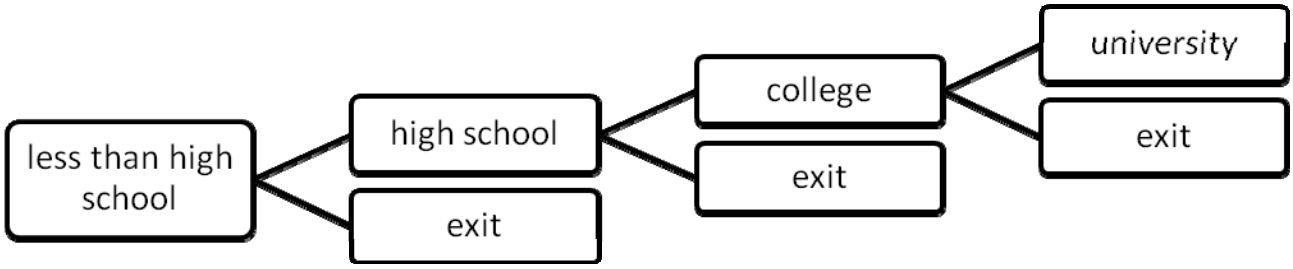

Table A1: Older youth cohort sample size over cycles 1 to 7, and respondents by age and regions in cycle 6 of the NLSCY

Older youth cohort sample size over cycles 1 to 6				
Cycle and age	Un-weighted Sample		Weighted Sample[3]	
C7 (2006-2007) 18-23 years[1]	6,501		NA	
C6 (2004-2005)	2,982[2]		1,509,944	
18	815		373,888	
19	758		385,038	
20	746		379,046	
21	663		371,972	
C5 (2002-2003)	4,424			
16	1,122			
17	1,052			
18	1,172			
19	1,078			
C4 (2000-2001)	4,506			
14	1,151			
15	1,101			
16	1,173			
17	1,081			
C3 (1998-1999)	4,317			
12	1,264			
13	875			
14	1,262			
15	916			
C2 (1996-1997)	4,519			
10	1,193			
11	1,056			
12	1,202			
13	1,068			
C1 (1994-1995)	4,519			
8	1,193			
9	1,056			
10	1,202			
11	1,068			
Respondents for:	2,982 (cycle 6)			
6 cycles	2,414		1,200,277 (funnel weight)[4]	
5 cycles	2,848			
4 cycles	2,962			
3 cycles	2,980			
Cycle and age	Canada	Ontario	Québec	Rest of Canada
C6 (2004-2005)	2,982	730	531	1,721
18	815	218	137	460
19	758	178	140	440
20	746	183	137	426
21	663	151	117	395

Sources: NLSCY's User Guide, C1 to C6; and author's calculation.

Notes: 1. NLSCY Consultation Guide 2008, HRSD Canada, April 2008. 2. From the 4,695 longitudinal respondents in the 16-21 years old youth cohort and after excluding 128 respondents from the cohort classified as age unknown (dead or moved to another country). 3. Longitudinal weight of cycle 6. 4. Funnel weight of cycle 6.

Table A2: Variables available in the NLSCY, retained for the data set by cycle and age of youth

1. Family background-Parent(s) Reported	
1. PMK education levels C1-C4	√
2. Family status C1-C5	√
3. Family "separation"- constructed from C2-C5	√
4. Province of residence C1-C6	√
5. Family income C1-C4	√
6. Number of siblings C1-C5	√
7. PMK not born in Canada C1-C5	NS
2. Community-Parent(s) Reported	
1. Community size C1-C5	√
2. Neighbourhood safety score C1	NS
3 Neighbourhood problems score C1	NS
4. Neighbours score C1	NS
5 Neighbours score C3 (if <16 years)	NS
3. Youth	
1. Ages (18-21) C6	√
2. Gender	√
3. Math scores 7-15 years C1-C4	√
4. Reading scores 7-15 years C2+C3	√
5. Cognitive measure (reading comprehension, problem-solving and decision-making) scores 16-17 years C4-C5	NS
6. Literacy assessment 18-19 years C6	
7. Numeracy assessment 20-21 years C6	
4. Child or Parenting Scores on Parent(s) Reported Scales	
1. Family functioning ("dysfunctional") 8-17 years C1-C4	√
2. Problem resolution 14-15 years C3-C4	NS
3. Behaviours 8-11 years C1-C2: hyperactive-inattentive√ /conduct disorders-physical aggression √/	√
pro-social NS/anxiety and emotional disorder NS/ indirect aggression NS/property offences NS	NS
4. Health status of child (excellent-very good/others) 8-17 years C1-C4	NS
5. Parenting 8-11 years C1-C2: positive interactions NS/ ineffective NS /consistent√ /rational NS	NS
5. Scores on Youth Reported Scales	
1. Health status (excellent-very good/others) 18-21 years C5-C6	√
2. Emotional Quotient: 4 and 5 dimensions (intrapersonal, interpersonal, adaptability, stress management, and general mood) C5-C6	NS
3. General Self Score (age 10 to 19 years) C3-C5	√
4. Parental Nurturance Score 10-15 years C3-C4	√
5. Parental Rejection Score 10-15 years C2-C4	NS
6. Parental Monitoring 10-15 years C3-C4	NS
7. Conflict Resolution – Mother/Father 16-17 years C4	NS
8. Pro-social 10-15 years C3-C4	NS
9. Neighbourhood structure score if 16-17 years C4-C5	√
10. Depression Scale (age 16 to 17) C4-C5	NS
11. Friends Score (age 10 to 17 years)	
12. Indirect Aggression Score (age 10 to 15 years)	NS
13. Anxiety and Emotional Disorder Score (age 10 to 15 years)	NS
14. Conduct Disorder/Physical Aggression Score (age 10 to 15 years)	√
15. Hyperactivity/Inattention Score (age 10 to 15 years)	√
16. Property Offence Score (age 10 to 15 years)	NS
6. Principals and Teachers: School quality index 7-15 years C1-C4 (from Lefebvre et al. 2008a)	√

Note: 1. Variables used (√), and in general not significant (NS).

Table A3: OLS estimation of obtaining high-school diploma, 18 to 21 year olds women and men, NLSCY, Québec 2004-2005

Variables	Specifications					
	(1)		(2)		(3)	
Age19	0.0806	(0.0666)	0.0243	(0.0655)	0.0139	(0.0643)
Age 20	0.199***	(0.0689)	0.125*	(0.0677)	0.0864	(0.0685)
Age 21	0.174**	(0.0717)	0.0689	(0.0685)	0.0744	(0.0720)
Female	0.246***	(0.0484)	0.172***	(0.0441)	0.118***	(0.0451)
PmkEduPrim	-0.138	(0.244)	-0.131	(0.220)	-0.100	(0.243)
PmkEduSec	0.145	(0.0909)	0.172**	(0.0809)	0.137	(0.0855)
PmkEduSomePSE	0.131	(0.0835)	0.127	(0.0784)	0.126	(0.0771)
PmkEduCol	0.185**	(0.0806)	0.166**	(0.0726)	0.135*	(0.0714)
PmkEduUni	0.154*	(0.0836)	0.118	(0.0773)	0.103	(0.0761)
TwoBioParents	0.0623	(0.0781)	0.0377	(0.0729)	0.0546	(0.0788)
TwoParents	0.0522	(0.107)	0.00994	(0.0970)	0.129	(0.0967)
Separation	0.0351	(0.0666)	0.0294	(0.0558)	-0.0200	(0.0584)
PmkImmig	0.0453	(0.109)	0.108	(0.0941)	0.0550	(0.0925)
Sibling1	-0.0210	(0.0706)	-0.00867	(0.0616)	0.0697	(0.0605)
Siblings2	-0.00662	(0.0673)	-0.00473	(0.0616)	0.0677	(0.0642)
ComSizeRural	-0.0502	(0.0721)	-0.0714	(0.0671)	-0.0997	(0.0656)
ComSize2	0.0905	(0.0586)	0.0622	(0.0568)	0.0557	(0.0574)
SchoolQIndex	-0.00567	(0.0180)	-0.0120	(0.0179)	-0.0291	(0.0183)
Q2FamInc	0.0844	(0.0718)	0.113*	(0.0660)	0.134**	(0.0654)
Q3FamInc	0.150**	(0.0721)	0.156**	(0.0691)	0.189***	(0.0698)
Q4FamInc	0.252***	(0.0894)	0.227***	(0.0825)	0.222***	(0.0806)
Q2Math	0.133	(0.166)	0.156	(0.154)	0.227	(0.168)
Q3Math	0.199	(0.149)	0.177	(0.141)	0.172	(0.159)
Q4Math	0.256*	(0.150)	0.235*	(0.141)	0.244	(0.157)
Q2Reading	0.221**	(0.0957)	0.204**	(0.0864)	0.267***	(0.0915)
Q3Reading	0.258***	(0.0901)	0.187**	(0.0836)	0.216**	(0.0886)
Q4reading	0.244**	(0.0993)	0.234**	(0.0917)	0.267***	(0.0954)
HealthE&VG	0.0716	(0.0521)	0.0266	(0.0471)	0.0645	(0.0497)
SelfEsteem	0.0267**	(0.0129)	0.0188*	(0.0112)	0.0119	(0.0133)
HyperActInat			-0.0388***	(0.00793)	-0.0393***	(0.00805)
Aggression11/15			0.0157	(0.0160)	0.0148	(0.0167)
Nurturance					-0.00133	(0.00514)
Constant	-0.541*	(0.299)	-0.0314	(0.279)	0.0512	(0.310)
Observations	423		422		366	
Adj. R^2	0.280		0.352		0.369	

Notes: 1. Bootstrapped standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: OLS estimation of post-secondary education participation if high-school diploma, 18 to 21 year olds women and men, NLSCY, Québec 2004-2005

Variables	Specifications					
	(1)		(2)		(3)	
Age19	0.0565*	(0.0342)	0.0611*	(0.0352)	0.0305	(0.0344)
Age 20	-0.0916	(0.0683)	-0.0816	(0.0618)	-0.0944	(0.0670)
Age 21	0.0375	(0.0413)	0.0522	(0.0490)	0.0101	(0.0454)
Female	0.0872**	(0.0438)	0.0849*	(0.0436)	0.0557	(0.0452)
PmkEduPrim	-0.392	(0.260)	-0.395	(0.265)	-0.504**	(0.228)
PmkEduSec	0.0277	(0.0863)	0.0297	(0.0874)	0.140**	(0.0664)
PmkEduSomePSE	0.130*	(0.0684)	0.134*	(0.0694)	0.136*	(0.0781)
PmkEduCol	0.116	(0.0705)	0.130*	(0.0741)	0.151*	(0.0772)
PmkEduUni	0.0931	(0.0773)	0.0958	(0.0761)	0.0842	(0.0894)
TwoBioParents	-0.0423	(0.0663)	-0.0474	(0.0660)	-0.0969*	(0.0504)
TwoParents	-0.0235	(0.0767)	-0.0194	(0.0784)	-0.0520	(0.0565)
Separation	0.0460	(0.0394)	0.0402	(0.0391)	0.0634*	(0.0354)
PmkImmig	-0.0552	(0.116)	-0.0670	(0.114)	-0.132	(0.119)
Sibling1	0.0311	(0.0483)	0.0345	(0.0472)	0.0213	(0.0386)
Siblings2	-0.00670	(0.0569)	0.000452	(0.0566)	0.00855	(0.0497)
ComSizeRural	0.0652	(0.0606)	0.0684	(0.0606)	0.0209	(0.0682)
ComSize2	0.0574	(0.0479)	0.0519	(0.0462)	0.0279	(0.0519)
SchoolQIndex	0.00213	(0.0105)	0.00310	(0.0111)	0.00240	(0.0120)
Q2FamInc	0.00730	(0.0427)	0.00397	(0.0421)	-0.00945	(0.0429)
Q3FamInc	0.0378	(0.0463)	0.0284	(0.0474)	0.00366	(0.0456)
Q4FamInc	0.0426	(0.0593)	0.0464	(0.0589)	0.0327	(0.0559)
Q2Math	0.0944	(0.169)	0.110	(0.177)	-0.111	(0.152)
Q3Math	0.105	(0.157)	0.120	(0.165)	-0.0637	(0.149)
Q4Math	0.159	(0.164)	0.179	(0.172)	-0.0430	(0.157)
Q2Reading	0.0603	(0.0632)	0.0656	(0.0630)	0.0427	(0.0634)
Q3Reading	-0.0511	(0.0806)	-0.0439	(0.0769)	-0.00963	(0.0710)
Q4reading	0.0731	(0.0633)	0.0677	(0.0641)	0.0959	(0.0637)
HealthE&VG	0.0924*	(0.0494)	0.0943*	(0.0496)	0.0405	(0.0468)
SelfEsteem	-0.00411	(0.0103)	-0.00338	(0.00987)	-0.00431	(0.0112)
HyperActInat			0.00590	(0.00808)	-0.00143	(0.00783)
Aggression11/15			-0.0180	(0.0149)	-0.0105	(0.0160)
Nurturance					0.00354	(0.00342)
Constant	0.614***	(0.225)	0.566**	(0.234)	0.855***	(0.206)
Observations	318		317		278	
Adj. R^2	0.210		0.213		0.223	

Notes: 1. Bootstrapped standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Quartiles of reading scores by quartiles of math scores in percentage and number of 18 to 21 year olds, NLSCY, Canada, 2004-2005

Quartile of math C1-C4, 7-15 year-olds	Quartiles of reading score C2-3, 10-15 year-olds					
	NA	1	2	3	4	Total
NA	100 125,012					100 125,012
1	10 34,220	42 144,179	28 96,379	14 48,480	7 23,111	100 346,369
2	8 27,522	25 88,670	27 93,553	23 80,685	17 58,071	100 348,501
3	7 22,565	16 54,532	23 78,266	31 105,355	25 84,749	100 345,467
4	9 30,862	9 29,389	16 55,314	23 80,666	43 148,364	100 344,595
Total	16 235,868	21 321,083	21 323,512	21 315,186	21 314,295	100 1,509,944

Source: author's calculation from the NLSCY's cycles 1 to 4.
Note: weighted estimation and rounding of percentage.

Table A6: Educational status by math and reading quartiles in percentage, 18 to 21 year olds, NLSCY, Canada, 2004-2005

	High-school diploma			Postsecondary participation			
Math Quartiles	**No**	**Yes**	**Total**	**No**	**College**	**University**	**Total**
NA	31	69	100	44	25	31	100
1	34	66	100	53	32	15	100
2	25	75	100	44	32	24	100
3	17	83	100	31	33	37	100
4	12	88	100	21	31	48	100
Total	23	77	100	38	31	31	100
	High-school diploma			**Postsecondary participation**			
Reading Quartiles	**No**	**Yes**	**Total**	**No**	**College**	**University**	**Total**
NA	28	72	100	45	29	26	100
1	34	66	100	49	36	15	100
2	24	76	100	41	33	27	100
3	15	85	100	30	31	39	100
4	14	86	100	27	28	45	100
Total	23	77	100	38	31	31	100

Source: author's calculation from the NLSCY's cycles 1 to 4.
Note: weighted estimation and rounding of percentage.

Table A7: Coefficients (p-value of bootstrapped standard errors) of multinomial estimation of schooling outcomes, 18 to 21 year olds women and men, NLSCY, Canada less Québec, 2004-2005

Variables	Women				Men			
	Outcomes				Outcomes			
	(2) College		(3) University		(2) College		(3) University	
Age19	0.281	(0.608)	0.276	(0.629)	0.744	(0.125)	-0.191	(0.753)
Age 20	1.413**	(0.021)	0.987*	(0.088)	0.651	(0.234)	0.176	(0.789)
Age 21	1.402**	(0.023)	1.075*	(0.070)	1.217**	(0.018)	0.744	(0.253)
Ont	0.608	(0.240)	-0.617	(0.203)	0.074	(0.889)	-0.113	(0.851)
Prai	-0.209	(0.684)	-1.109**	(0.035)	-0.416	(0.485)	-0.513	(0.452)
Alb	0.775	(0.181)	-1.763**	(0.017)	-0.273	(0.625)	-1.301**	(0.042)
BC	0.516	(0.452)	-0.795	(0.211)	0.042	(0.951)	-1.679**	(0.047)
PmkEduPrim	-1.868	(0.854)	-1.127	(0.837)	-1.443	(0.760)	-1.769	(0.846)
PmkEduSec	0.034	(0.958)	-0.328	(0.662)	-1.170*	(0.083)	-1.748**	(0.049)
PmkEduSomePSE	0.104	(0.821)	0.114	(0.812)	-0.638	(0.234)	-0.739	(0.260)
PmkEduCol	0.340	(0.550)	0.417	(0.506)	-0.436	(0.421)	-0.092	(0.892)
PmkEduUni	0.298	(0.830)	1.729	(0.214)	-0.265	(0.730)	0.913	(0.233)
TwoBioParents	0.612	(0.305)	1.705***	(0.008)	0.190	(0.774)	-0.144	(0.854)
TwoParents	0.974	(0.249)	2.656***	(0.003)	-0.334	(0.705)	-0.141	(0.901)
Separation	-0.496	(0.364)	-1.077*	(0.075)	-0.020	(0.966)	-1.013	(0.101)
PmkImmig	-0.547	(0.418)	0.534	(0.413)	0.336	(0.601)	0.286	(0.665)
Sibling1	-0.053	(0.931)	-0.309	(0.640)	-0.306	(0.631)	0.015	(0.983)
Siblings2	0.081	(0.901)	0.163	(0.810)	-0.457	(0.476)	-0.750	(0.316)
ComSizeRural	-0.075	(0.867)	-0.742	(0.108)	-0.008	(0.984)	-1.565***	(0.005)
ComSize2	-0.284	(0.521)	-0.625	(0.139)	-0.333	(0.439)	-0.169	(0.745)
SchoolQIndex	-0.116	(0.393)	0.069	(0.633)	0.129	(0.296)	-0.130	(0.375)
Q2FamInc	0.425	(0.393)	-0.045	(0.933)	-0.073	(0.893)	-0.444	(0.541)
Q3FamInc	0.362	(0.535)	-0.237	(0.698)	0.153	(0.759)	-0.001	(0.998)
Q4FamInc	0.150	(0.814)	-0.098	(0.878)	0.317	(0.594)	-0.111	(0.884)
Q2Math	0.018	(0.967)	0.692	(0.165)	0.070	(0.889)	1.680	(0.126)
Q3Math	-1.029**	(0.045)	0.093	(0.854)	0.025	(0.960)	2.001*	(0.055)
Q4Math	1.729	(0.539)	3.483	(0.210)	0.378	(0.513)	2.853***	(0.008)
Q2Reading	-0.158	(0.763)	0.604	(0.283)	0.114	(0.810)	0.185	(0.817)
Q3Reading	0.008	(0.987)	0.557	(0.331)	-0.820*	(0.095)	0.231	(0.728)
Q4reading	-0.037	(0.954)	0.860	(0.152)	-0.235	(0.649)	0.263	(0.681)
HealthE&VG	0.084	(0.818)	1.054***	(0.008)	0.137	(0.720)	0.573	(0.225)
SelfEsteem	0.058	(0.535)	0.023	(0.810)	-0.015	(0.891)	-0.053	(0.698)
HyperActInat	-0.043	(0.463)	-0.093	(0.136)	-0.095*	(0.098)	-0.174**	(0.021)
Aggression11/15	-0.110	(0.469)	-0.421**	(0.023)	-0.095	(0.432)	0.001	(0.997)
Nurturance	-0.013	(0.753)	0.022	(0.611)	0.039	(0.303)	0.057	(0.301)
Constant	-0.734	(0.669)	-2.162	(0.252)	-0.124	(0.948)	0.258	(0.913)
Observations	809				664			

1. Base outcome is high-school (HS) graduate and non-PSE participant. P-values in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Coefficients (p-value of bootstrapped standard errors) of multinomial estimation of four schooling outcomes, 18 to 21 year olds women, NLSCY, Canada, 2004-2005

Variables	Outcomes					
	(1) High-school continuer		(2) High-school graduate and non-PSE participant		(3) PSE participant	
Age19	-2.991***	(0.000)	-1.401*	(0.060)	-1.183*	(0.074)
Age 20	-2.046	(0.227)	-0.015	(0.987)	0.792	(0.345)
Age 21	-3.498	(0.792)	-1.119	(0.181)	0.106	(0.891)
Que	-1.771	(0.101)	-2.025	(0.661)	-0.809	(0.459)
Ont	-0.703	(0.502)	-1.091	(0.288)	-0.856	(0.406)
Prai	-0.130	(0.943)	0.559	(0.760)	0.152	(0.932)
Alb	-1.378	(0.709)	-0.716	(0.534)	-0.793	(0.483)
BC	-0.442	(0.905)	-0.087	(0.981)	-0.235	(0.948)
PmkEduPrim	-1.510	(0.639)	-0.825	(0.754)	-2.554	(0.233)
PmkEduSec	0.410	(0.661)	0.250	(0.767)	0.775	(0.286)
PmkEduSomePSE	0.174	(0.828)	0.008	(0.991)	0.562	(0.363)
PmkEduCol	-0.166	(0.862)	0.392	(0.666)	1.250	(0.133)
PmkEduUni	0.098	(0.994)	0.763	(0.944)	2.212	(0.837)
TwoBioParents	0.518	(0.509)	0.598	(0.461)	1.015	(0.149)
TwoParents	-0.034	(0.974)	-0.539	(0.641)	0.851	(0.394)
Separation	-0.312	(0.699)	0.424	(0.571)	0.050	(0.937)
PmkImmig	0.392	(0.914)	0.720	(0.774)	0.728	(0.770)
Sibling1	0.822	(0.375)	1.542*	(0.062)	1.277*	(0.089)
Siblings2	0.495	(0.568)	0.904	(0.286)	0.801	(0.292)
ComSizeRural	0.129	(0.869)	0.407	(0.579)	0.357	(0.574)
ComSize2	0.854	(0.294)	0.455	(0.553)	0.286	(0.683)
SchoolQIndex	-0.055	(0.819)	0.143	(0.447)	0.107	(0.539)
Q2FamInc	0.348	(0.654)	0.592	(0.428)	0.722	(0.283)
Q3FamInc	2.279	(0.107)	1.942	(0.163)	2.251*	(0.090)
Q4FamInc	0.386	(0.747)	0.491	(0.666)	0.760	(0.474)
Q2Math	0.559	(0.494)	0.321	(0.667)	0.451	(0.535)
Q3Math	1.377	(0.181)	1.246	(0.145)	0.971	(0.214)
Q4Math	1.711	(0.472)	0.209	(0.940)	1.357	(0.551)
Q2Reading	0.032	(0.968)	0.064	(0.935)	0.618	(0.442)
Q3Reading	0.142	(0.869)	0.014	(0.985)	0.425	(0.560)
Q4reading	-0.659	(0.487)	-0.947	(0.264)	0.085	(0.909)
HealthE&VG	0.269	(0.663)	0.421	(0.473)	1.115**	(0.044)
SelfEsteem	-0.066	(0.683)	0.054	(0.675)	-0.011	(0.923)
HyperActInat	0.158*	(0.076)	0.048	(0.530)	-0.021	(0.770)
Aggression15	-0.196	(0.357)	-0.228	(0.116)	-0.433***	(0.004)
Nurturance	0.148**	(0.039)	0.057	(0.264)	0.059	(0.244)
Constant	-2.418	(0.334)	-3.067	(0.141)	-2.402	(0.213)
Observations	1,168					

Notes: 1. Base outcome is high-school (HS) drop-out. P values in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Coefficients (p-value of bootstrapped standard errors) of multinomial estimation of three schooling outcomes, 18 to 21 year olds men, NLSCY, Canada, 2004-2005

Variables	Outcomes					
	(1) High-school continuer		(2) High-school graduate and non-PSE participant		(3) PSE participant	
Age19	-1.328	(0.107)	0.214	(0.738)	0.566	(0.384)
Age 20	-1.507	(0.704)	0.911	(0.285)	1.124	(0.180)
Age 21	-3.092	(0.759)	0.104	(0.889)	1.123	(0.110)
Que	-1.013	(0.338)	-2.703***	(0.003)	-1.889**	(0.020)
Ont	0.364	(0.685)	-0.548	(0.473)	-0.517	(0.504)
Prai	-0.926	(0.340)	-0.689	(0.367)	-1.080	(0.154)
Alb	-0.195	(0.904)	-0.134	(0.926)	-0.783	(0.583)
BC	-1.766	(0.694)	-0.332	(0.939)	-0.830	(0.848)
PmkEduPrim	-1.642	(0.667)	-0.109	(0.931)	-1.031	(0.435)
PmkEduSec	-1.239	(0.200)	0.316	(0.733)	-0.727	(0.397)
PmkEduSomePSE	-0.443	(0.637)	0.607	(0.503)	0.145	(0.869)
PmkEduCol	0.961	(0.302)	0.948	(0.282)	0.906	(0.281)
PmkEduUni	1.570	(0.326)	1.246	(0.413)	1.327	(0.374)
TwoBioParents	0.530	(0.537)	1.859***	(0.008)	1.802***	(0.008)
TwoParents	-1.022	(0.378)	0.950	(0.313)	0.629	(0.493)
Separation	0.427	(0.594)	0.757	(0.292)	0.481	(0.485)
PmkImmig	-1.283	(0.269)	-1.112	(0.260)	-1.201	(0.208)
Sibling1	-2.211**	(0.014)	-1.763**	(0.032)	-1.679**	(0.033)
Siblings2	-1.948**	(0.037)	-0.962	(0.247)	-1.303*	(0.100)
ComSizeRural	-0.497	(0.544)	-0.532	(0.421)	-1.091*	(0.079)
ComSize2	0.858	(0.242)	0.702	(0.307)	0.375	(0.565)
SchoolQIndex	-0.025	(0.912)	-0.127	(0.501)	-0.078	(0.680)
Q2FamInc	1.281	(0.121)	0.870	(0.232)	0.465	(0.503)
Q3FamInc	0.578	(0.512)	0.628	(0.429)	0.568	(0.458)
Q4FamInc	0.600	(0.691)	0.693	(0.633)	0.675	(0.636)
Q2Math	0.154	(0.847)	0.205	(0.759)	0.500	(0.465)
Q3Math	-0.319	(0.715)	-0.093	(0.907)	0.496	(0.529)
Q4Math	0.789	(0.627)	1.259	(0.408)	2.449*	(0.099)
Q2Reading	0.011	(0.990)	0.281	(0.715)	0.516	(0.497)
Q3Reading	-0.729	(0.373)	-0.185	(0.803)	-0.547	(0.443)
Q4reading	-0.784	(0.509)	-0.017	(0.988)	0.081	(0.941)
HealthE&VG	0.117	(0.857)	0.193	(0.761)	0.427	(0.505)
SelfEsteem	-0.025	(0.875)	-0.054	(0.658)	-0.070	(0.573)
HyperActInat	-0.066	(0.432)	-0.107	(0.177)	-0.185***	(0.010)
Aggression11	0.072	(0.676)	-0.088	(0.578)	-0.147	(0.345)
Nurturance	0.076	(0.199)	0.059	(0.235)	0.087*	(0.076)
Constant	2.073	(0.439)	0.819	(0.752)	1.785	(0.484)
Observations	1,010					

Notes: 1. Base outcome is high-school (HS) drop-out. P values in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Bibliography

Blanden, J. and P. Gregg (2004), "Family Income and Educational Attainment: a Review of Approaches and Evidence for Britain," *Oxford Review of Economic Policy*, XX, 245-263.

Blanden, J and S. Machin (2004), "Educational Inequality and the Expansion of UK Higher Education," *Scottish Journal of Political Economy*, 51(2): 230-249.

Belzil, C. and J. Hansen (2006), "Educational Attainment in Canada: Effects of Individual Attributes and Expected Outcomes," Human Resources and Social Development Canada, Industry Canada, Social Sciences and Humanities Research Council, Working Paper 2006 C-11.

Cameron, S. and, J. Heckman (1998), "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy*, CVI, 262-333.

Cameron, S. and J. Heckman (2001), "The Dynamics of Educational Attainment for Black, Hispanic and White Males," *Journal of Political Economy*, CIX (2001), 455-499.

Carneiro, P. and J. Heckman (2002), "The Evidence on Credit Constraints in Post-Secondary Schooling," *The Economic Journal*, 112(482): 705–734.

Carneiro, P. and J. Heckman (2003), Human Capital Policy, In J. J. Heckman, A. B. Krueger, and B. M. Friedman (Eds.), Inequality in America: What Role for Human Capital Policies? Cambridge, MA: MIT Press.

Christofides, L., J. Cirello and M. Hoy (2001), "Family Income and Post-secondary Education in Canada," *The Canadian Journal of Higher Education* 31(1), 177-208

Coelli, M. (2005), "Parental Income Shocks and the Education Attendance of Youth," Mimeo, Department of Economics, University of British Columbia.

Corak, M., G. Lipps et J. Zhao (2003), *Revenu familial et participation aux études postsecondaires*. Direction des études analytiques : documents de recherche. No 11F0019MIF2003210 au catalogue. Ottawa : Statistique Canada.

Cunha, F. (2007), "A Time to Plant and a Time to Reap," Mimeo, Department of Economics, University of Chicago.

Cunha, F. and J. Heckman (2007), "Formulating, Identifying and Estimating the Technology of Cognitive and Non-cognitive Skill Formation," University of Chicago, Department of Economics, Forthcoming, *Journal of Human Resources*.

Cunha F. and J. Heckman (2006), "Investing in our Young People," mimeo, Department of Economics, University of Chicago.

Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov (2006), "Interpreting Evidence of Life Cycle Skill Formation," in *Handbook of the Economics of Education*, Vol. 1. Ed. Eric A. Hanushek and Finis Welch, New York: Elsevier, 697-812.

Cunha, Flavio and James J. Heckman (2007), "The Technology of Skill Formation," *American Economic Review, Papers and Proceedings* 97(2), 31-47.

Drolet, M. (2005), "Participation aux études postsecondaires au Canada : le rôle du revenu et du niveau de scolarité des parents a-t-il évolué au cours des années 1990?" Direction des études analytiques: documents de recherche. No 11F0019MIF2005243 au catalogue. Ottawa: Statistique Canada.

Dynarski, Susan (2003), "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion," *American Economic Review* 93:1, 279-288.

Ermisch, John, and Marco Francesconi (2001), "Family Matters: Impacts of Family Background on Educational Attainments," *Economica* 68(270), 137-56.

Ellwood, D. and T. Kane (2000), "Who is Getting a College Education? Family Background and the Growing Gaps in Enrollment," in Sheldon Danziger and Jane Waldfogel, eds., *Securing the Future*. New York: Russell Sage.

Finnie, R and R. Mueller (2007), "The Effects of Family Income, Parental Education and Other Background Factors on Access to Post-Secondary Education in Canada: Evidence from the YITS," Forthcoming MESA Working Paper.

Finnie, R, A. Sweetman and Eric Lascelles (2004), "Who Goes: The Direct and Indirect Effects of Family Background on Access to Post-Secondary Education" (with), in C. Beach, R. Boadway, and M. McInnis (eds.), *Higher Education in Canada*, Kingston: John Deutsch Institute, McGill-Queen's University Press, pp. 295-338.

Fontaine, N., Carbonneau, R., Barker, E. D., Vitaro, F., Hébert, M., Côté, S. M., Nagin, D. S., Zoccolillo, M. and Tremblay, R. E. (2008), "Females' Hyperactivity and Physical Aggression During Childhood Predict Adjustment Problems in Early Adulthood: A 15-year Longitudinal Study,". *Archives of General Psychiatry, 65*(3): 320-328.

Frenette, M. (2004), "Access to College and University: Does Distance Matter?" *Canadian Public Policy – Analyse de politiques*, 30(4): 427–443.

Frenette, Marc (2007), "Why are Youth from Lower-income Families Less Likely to Attend University? Evidence from Academic Abilities, Parental Influences, and Financial Constraints," Statistics Canada, Analytical Studies Research Paper Series No. 295.

Hango, Darcy and Patrice de Broucker (2007), "Education-to-Labour Market Pathways of Canadian Youth: Findings from the Youth in Transition Survey," Statistics Canada Catalogue no. 81-595-MIE – No. 054.

Heckman, James J. (2000), "Policies to Foster Human Capital," *Research in Economics* 54(1), 3- 56.

Heckman, James J. (2007), "The Economics, Technology and Neuroscience of Human Capability Formation," NBER Working Paper No. 13195.

Heckman, James J., and Dimitry V. Masterov (2007), "The Productivity Argument for Investing in Young Children," NBER Working Paper No. 13016.

Heckman, J., J. Stixrud and S. Urzua. (2006), "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," NBER Working Paper 12006.

Hansen, J. (2007), "Education and Early Labour Market Outcomes in Canada," Report prepared for Prepared for Learning Policy Directorate, Strategic Policy, Human Resources and Social Development Canada, SP-793-12-07E.

Hansen, J. (2006), "Returns to University Level Education: Variations Within Disciplines, Occupations and Employment Sectors," Report prepared for Prepared for Learning Policy Directorate, Strategic Policy, Human Resources and Social Development Canada, SP-662-09-06E.

Keane, M. and K. Wolpin (1997), "The Career Decisions of young Men," *The Journal of Political Economy* 105(3): 473-522.

Keane, M. and K. Wolpin (2001), "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment" *International Economic Review* 42(4), 1051-1103.

Leblanc, N., Boivin, M., Dionne, G., Brendgen, M., Vitaro, F., Tremblay, R. E. and Pérusse, D. (2008), "The Development of Hyperactive/impulsive Behaviors during the Preschool Years: The Predictive Validity of Parental Assessments," Forthcoming, *Journal of Abnormal Child Psychology*. [Online first 2008 March 11]

Lefebvre, Pierre, Philip Merrigan and Matthieu Verstraete (2008a), "The Effects of Schools and Family Functioning on Math Scores: A Canadian Longitudinal Analysis," Working Paper, UQAM, presented at the Annual Meeting of the Canadian Economics Association, Vancouver, June 2008.

Lefebvre, Pierre, Philip Merrigan and Matthieu Verstraete (2008b), **"**Private Schools Do Make a Difference on Math Scores: Canadian Longitudinal Evidence using the Primary to Secondary School Transitions as a Natural Experiment," Working paper, UQAM, August.

Lillard L, Willis R (1994), "Intergenerational Educational Mobility: Effects of Family and State in Malaysia," *Journal of Human Resources* 29(4):1126–1167.

Rahman, Atiq, Jerry Situ and Vicki Jimmo (2005), "Participation in Postsecondary Education: Evidence from the Survey of Labour Income Dynamics," Statistics Canada Catalogue no. 81-595-MIE2005036.

Rivard, Maud and M. Raymond (2004), "The Effect of Tuition Fees on Post-secondary Education in Canada in the late 1990s," Working Paper 2004-09, Department of Finance, Ottawa.

Shaienks, Danielle and Tomasz Gluszynski (2007), "Participation in Postsecondary Education: graduates, continuers and Drop Outs, Results from YTIS Cycle 4," Statistics Canada Catalogue no. 81-595-MIE – No. 059.

Shaienks, Danielle,Judy Eisl-Culkin and Patrick Bussière (2006), "*Follow-Up on Education and Labour Market Pathways of Young Canadians Aged 18 to 20 – Results from YITS Cycle 3*," Ottawa: Human Resources and Social Development Canada and Statistics Canada, Statistics Canada Catalogue no. 81-595-MIE – No. 045.

Todd, Petra and Kenneth Wolpin (2007), "The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps," *Journal of Human Capital*, 1(1): 91-136.

Tremblay, R. E., Larocque, D., Boivin, M., Pérusse, D., Zoccolillo, M. and Pihl, R. O. (2008), "Physical Aggression During Infancy and Onset of Male Conduct Disorder," in W. Koops, N. W. Slot and R. Loeber (Eds.), *A Developmental Approach of Antisocial Behavior*, Forthcoming, Abingdon, United Kingdom : Psychology Press.