



# **Parental Education and Postsecondary Attainment: Does the Apple Fall Far from the Tree?**

## **Appendix**

**Ken Chatoor, Emily MacKay and Lauren Hudak**  
Higher Education Quality Council of Ontario

Published by

## Higher Education Quality Council of Ontario

1 Yonge Street, Suite 2402  
Toronto, ON Canada, M5E 1E5

Phone: (416) 212-3893  
Fax: (416) 212-3899  
Web: [www.heqco.ca](http://www.heqco.ca)  
E-mail: [info@heqco.ca](mailto:info@heqco.ca)

### **Cite this publication in the following format:**

Chatoor, K., MacKay, E. & Hudak, L. (2019). *Parental Education and Postsecondary Attainment: Does the Apple Fall Far from the Tree? Appendix* Toronto: Higher Education Quality Council of Ontario.

© Queen's Printer for Ontario, 2019

# Appendix A

## 1. Data Analysis

The data analysis was carried out using STATA 15 at the York University Research Data Centre. We used Statistics Canada’s Longitudinal and International Study of Adults (LISA) data set linked to administrative tax files, and the Programme for the International Assessment of Adult Competencies and Skills (PIAAC).

- **Longitudinal and International Study of Adults (Wave 2)** — LISA, a longitudinal survey that is conducted every two years, provides the most up-to-date data on income, education and family history. It collects information across Canada about individuals’ jobs, education, health and family. It is linked to administrative tax files, such as the T1 Family File (T1FF), which have been used to provide detailed information on income and income trends (Heisz, Notten & Situ, 2016). LISA is also linked to a subset of respondents to the PIAAC. The availability of T1FF forms allowed the creation of the Intergenerational Family File, which was linked to LISA Wave 2 (Hemeon, 2016; Xie & Rainville, 2016), and used in this report.
- **Programme for the International Assessment of Adult Competencies and Skills** — PIAAC (2012) is used to assess literacy, numeracy and problem-solving skills in technology-rich environments. It contains self-reported information on education, skill-use at work, labour market outcomes and parental education history. It also asks questions regarding employment activity and education.

All final results were weighted as per the LISA user guide. The linking of LISA to the Intergenerational Family File was done per Statistics Canada user guides. To account for the sample design of the survey, bootstrap weighting procedures with respondent survey weights (RPW) were employed for all analyses.

From a technical standpoint, the LISA Intergenerational Family File (LISA-IFF) was created by appending T1FF tax data files from 1982–2013 and linking them to the existing LISA Wave 2 file. Subsequently, all data in LISA Wave 2 was also in the LISA-IFF file. In order to assess intergenerational outcomes, a variable was derived that provided the parental level of income of each eligible survey respondent when the respondent was 19. Due to the specific years of tax data provided, analyses that include parental income are limited to those between the ages of 20 and 50. An age of 20 years old was chosen since this would allow us to include individuals who may have graduated from a one-year diploma program.

Regressions where income was the independent variable were done using log-log linear regressions. The natural log was used for individual income (independent variable) and parental income at age 19 (dependent variable) in all regression models.

Logistic regression analyses were used to determine the extent to which parental education affects the completion of any type of PSE credential (Turcotte, 2011). Logistic regression was used since the independent variable, “Completion of *any type* of PSE credential,” is a binary variable coded as “No PSE credential attainment” (0), and “PSE credential attained” (1). In the case of youth expectations, the independent variable is “Expect to complete *any type* of PSE credential,” which is coded as “Does not expect to complete PSE credential” (0), and “Expects to complete *any type* of PSE credential” (1). For all

job characteristics regressions, the independent variable is “Has job characteristic,” which is coded as “Does not have job characteristic” (0), and “Has job characteristic” (1).

We present all of our regression results as marginal effects. They are shown as percentages, which give an estimated probability effect that a factor in the model has on the outcomes we are testing. Each marginal effect is compared to the “base” of the factor. For example, the value for “college,” “university” or “graduate” is the probability effect that having a “college,” “university” or “graduate” credential has on an outcome compared to an individual who doesn’t have a PSE credential, since “no PSE credential” is the “base” of this factor. Another example would be for females, where the predicted marginal effect is the probability effect that being “female” has on an outcome compared to “males” since male is the “base” of this factor.

Our research questions that consider the highest level of PSE attainment follow the OECD International Standard Classification of Education levels. In these analyses, the independent variable is sorted by degree type. For example, when looking at youth expectations to complete PSE, the independent variable, “Highest Level of PSE Credential Expected to be Completed” was an ordered categorical variable with four coded values. This was coded as No PSE (0), College (1), University (2), and Graduate or professional degree programs (3). As this independent variable has four ordered categories, ordered logistic regression analyses were used. Those with missing household incomes or below \$500 were dropped from intergenerational analyses (Corak, Curtis & Phipps, 2010). The expectations to complete PSE were based on a LISA question whose sample universe only included individuals between the ages of 16 and 24 who were not enrolled in education. Income for both individuals and households were transformed for regression analysis to achieve normality in income distribution. A significance level of  $p < 0.05$  was applied in all cases.

## 2. Data Constrictions and Caveats

A lack of ideal data has been a major issue for conducting research on higher education. While some data does exist on intergenerational family income, K-12, higher education and tax administrative information, much of this data is separate and formatted in such a way that it is difficult to link. The LISA data set partially overcomes these issues. However, some of the work in this study is limited by the age, availability and extent of the information.

The Intergenerational Family File is limited by the year of available tax files, so our sample range by age was restricted to those between the ages of 20 and 50 for any analyses that included parental income at 19. As a result, we were not able to run regression models across the entire age cohort, which may have allowed for a more comprehensive analyses of long-term trends. Since higher education is highly influenced by provincial policy, we refrain from running regression analyses that explicitly compare Ontario to the rest of Canada. We instead run regressions nationally and in Ontario, and discuss these results in the context of Ontario policy.

As LISA is a largely self-reported survey, it is possible that there may be sample biases in the response of surveyed individuals. However, because LISA is linked to administrative data that provides information on immigration, geography, tax data, family structure, income and immigration, the risk of false responses is significantly lower for these variables. Lower income households and new immigrants historically have lower rates of returning taxes, which may skew the data toward higher income earners. This is a common issue for research that uses administrative tax data.

## Appendix B

### Regression Models

In Table B1 the dependent variable is “Completion of *any type* of PSE credential,” which is coded as “No PSE credential attainment” (0), and “PSE credential attained” (1).

**Table B1: Logistic Regression Output for PSE Credential Attainment, Canada (N=6,500)**

PSE Credential Attainment		Average Marginal Effects (%) (p value)
Family Income at Age 19		2.2% <sup>^</sup> <b>*(0.022)</b>
Non-First-Generation		19.3% <b>*(0.000)</b>
Female		9.0% <b>*(0.000)</b>
Household Size at Age 19		1.1% (0.123)
Geography: As compared to small population centres (<100,000)	Medium-sized Population Centre (100,000–500,000)	0.3% (0.877)
	Large Population Centre (>500,000)	7.8% <b>*(0.000)</b>
Immigrant		-4.5% (0.142)
Visible Minority		5.4% (0.101)
Indigenous		-13.2% <b>*(0.012)</b>
Age		0.7% <sup>^^</sup> <b>*(0.000)</b>

\* Significant at the 5% level (p<0.05)

<sup>^</sup> Marginal effect for income represents the probability effect for a 10% increase in income

<sup>^^</sup> Marginal effect for age represents probability effect for a single unit increase of age (years)

Source: LISA-IFF (2014)

In Tables B2 and B3 the dependent variable is “Expect to complete *any type* of PSE credential,” which is coded as “Does not expect to complete PSE credential” (0), and “Expects to complete *any type* of PSE credential” (1). Only individuals between the ages of 20 and 50 were used for this analysis.

**Table B2: Logistic Regression Output for PSE Credential Completion, Ontario (N=1,500)**

PSE Credential Attainment Ontario		Average Marginal Effects (%) (p value)
Family Income at Age 19		2.7% <sup>^</sup> (0.154)
Non-First-Generation		<b>19.5%</b> <b>*(0.000)</b>
Female		<b>9.6%</b> <b>*(0.003)</b>
Household Size at Age 19		0.1% (0.538)
Geography: As compared to small population centres (<100,000)	Medium-sized Population Centre (100,000–500,000)	-6.8% (0.249)
	Large Population Centre (>500,000)	6.6% (0.151)
Immigrant		1.0% (0.845)
Visible Minority		1.2% (0.812)
Indigenous		-27.2% (0.119)
Age		<b>0.7%<sup>^^</sup></b> <b>*(0.000)</b>

\* Significant at the 5% level ( $p < 0.05$ )

<sup>^</sup> Marginal effect for income represents the probability effects for a 10% increase in income

<sup>^^</sup> Marginal effect for age represents probability effect for a single unit of age (years)

Source: LISA-IFF (2014)

**Table B3: Logistic Regression Output for Effects of Sociodemographic Characteristics on Likelihood of Youth Expecting to Complete PSE Credential, Canada (N=3,000)**

Expect to Complete PSE Credential		Average Marginal Effects (%) (p value)
Family Income		<b>4.2%<sup>^</sup></b> <b>(0.002)</b>
Non-First-Generation		<b>11.0%</b> <b>*(0.000)</b>
Lone Parent Household		-2.3% (0.334)
Female		<b>5.8%</b> <b>*(0.00)</b>
Household Size		-0.6% (0.312)
Geography: As compared to small population centres (<100,000)	Medium-sized Population Centre (100,000–500,000)	2.9% (0.261)
	Large Population Centre (>500,000)	<b>5.3%</b> <b>*(0.029)</b>
Immigrant		-2.3% (0.479)
Visible Minority		4.6% (0.075)
Indigenous		-0.04% (0.934)

Source: LISA-IFF (2014)

<sup>^</sup> Marginal effect for income represents the probability effects for a 10% increase in income

<sup>^^</sup> Marginal effect for age represents probability effect for a single unit of age (years)

\* Significant at the 5% level (p<0.05)

**Table B4: Linear Regression Output for Effect of Sociodemographic Characteristics on Average Income Two and 10 Years after Graduation, Canada (N=3,000)**

Average Income		Two Years After Graduation	10 Years After Graduating
		Average Marginal Effects (%) (p value)	Average Marginal Effects (%) (p value)
PSE Credential Type	College	70.6% *(0.000)	30.8% *(0.000)
	University	83.0% *(0.000)	65.3% *(0.000)
	Graduate	123.8% *(0.000)	91.0% *(0.000)
Family Income at Age 19		8.8%^ *(0.012)	7.3%^ (0.029)
Non-First-Generation		-7.4% (0.419)	13.9% (0.112)
Female		-26.4% *(0.002)	-46.8% *(0.000)
Household Size at Age 19		-0.4% (0.892)	-3.7% (0.114)
Geography: As compared to small population centres (<100,000)	Medium-sized Population Centre (100,000–500,000)	-6.8% (0.748)	9.3% (0.171)
	Large Population Centre (>500,000)	3.3% (0.648)	-4.1% (0.568)
Immigrant		17.4% (0.672)	3.6% (0.778)
Visible Minority		-66.4% *(0.005)	-4.1% (0.737)
Indigenous		-35.4% (0.339)	14.0% (0.277)
Age at Graduation		3.8%^ *(0.000)	0.9%^ *(0.263)

\* Significant at the 5% level ( $p < 0.05$ )

^ Marginal effect for income represents the probability effects for a 10% increase in income

^^ Marginal effect for age represents probability effect for a single unit of age (years)

Source: LISA-IFF (2014)



**Table B5: Logistic Regression Output for Effect of Sociodemographic Factors on Job Characteristics, Canada (N=5,000)**

Job Market Characteristic		Pension	Work Group RRSP	Permanent Job	Manager Status	Bonus
		Average Marginal Effects (%) (p value)	Average Marginal Effects (%) (p value)	Average Marginal Effects (%) (p value)	Average Marginal Effects (%) (p value)	Average Marginal Effects (%) (p value)
PSE Credential Type	College	15.3% *(0.000)	7.2% *(0.000)	13.8% *(0.000)	1.4% (0.498)	2.2% (0.150)
	University	30.8% *(0.000)	11.5% *(0.000)	15.8% *(0.000)	7.7% (0.002)	8.0% *(0.008)
	Graduate	42.9% *(0.000)	14.5% *(0.000)	21.1% *(0.000)	16.0% (0.00)	0.5% (0.823)
Family Income at Age 19		-2.5% <sup>^</sup> *(0.034)	2.5% <sup>^</sup> (0.125)	1.1% <sup>^</sup> (0.310)	1.0% <sup>^</sup> (0.348)	-0.2% <sup>^</sup> (0.885)
Non-First-Generation		-3.5% (0.093)	2.0% (0.169)	-1.9% (0.314)	-2.5% (0.206)	2.3% (0.163)
Female		2.2% (0.238)	-7.5% *(0.000)	0.4% (0.826)	-10.8% *(0.000)	-7.1% (0.000)
Household Size at Age 19		0.3% (0.675)	-1.5% (0.134)	-1.6% (0.074)	-0.3% (0.676)	-1.1% (0.343)
Geography: As compared to small population centres (<100,000)	Medium-sized Population Centre (100,000–500,000)	1.4% (0.601)	-0.7% (0.782)	0.7% (0.801)	1.2% (0.632)	1.4% (0.486)
	Large Population Centre (>500,000)	0.7% (0.778)	-0.7% (0.785)	2.6% (0.294)	0.6% (0.815)	4.4% (0.029)
Immigrant		-15.9% *(0.000)	1.4% (0.723)	2.3% (0.525)	-5.0% (0.239)	3.2% (0.404)
Visible Minority		-10.9% *(0.005)	5.3% (0.139)	0.0% (0.990)	-10.3% *(0.016)	-0.3% (0.938)
Years since Graduation		0.7% <sup>^^</sup> *(0.000)	0.6% <sup>^^</sup> *(0.000)	1.1% <sup>^^</sup> *(0.000)	0.8% <sup>^^</sup> *(0.000)	0.4% <sup>^^</sup> (0.000)

\* Significant at the 5% level (p<0.05)

<sup>^</sup> Marginal effect for income represents the probability effects for a 10% increase in income

<sup>^^</sup> Marginal effect for years since graduation represents probability effect for a single unit increase (years)

Source: LISA-IFF (2014)