UNIVERSITY OF THE BASQUE COUNTRY - UPV/EHU

FACULTY OF ECONOMICS

Social class segregation in upper-secondary school choice in Canada

---

A thesis submitted by Pedro Iraburu for the

Master in Economics: Empirical Applications and Policies

Supervised by:

Petr Mariel

30th of July, 2020

Bilbo

CC BY NC ND

Abstract

The aim of this work is to analyze how the spoken language at home influences the parental decisions when selecting secondary schools in Canada. There is a vast literature on the factors affecting school choice, although very little research has been made on bilingual or multilingual countries. We take data from the 2012 PISA questionnaires to estimate the probability of choosing a determinate type of school, discriminating by instructional language and financing. The results show clearly that the impact of bilingualism dilutes otherwise very important socio-economic variables.

Keywords: School choice, bilingualism, multinomial logit.

Acknowledgments

INDEX

List of Figures

List of Tables

1 Introduction

The goal of the thesis is to cast light on the factors affecting the parental school choice in Canada, while taking into account the role of bilingualism. Canada is divided into 10 provinces and 3 territories. As these last are quite small, they were not included in the Programme for International Student Assessment (PISA) study (OECD, 2012), which we make use to extract the data necessary for our analysis.

There exists a wide array of languages coexisting in Canada, including many indigenous ones. In spite of this, only 209,570 individuals speak aboriginal languages in private (Statistics Canada, 2016) and thus, were not considered in our analysis. The two main languages in Canada are English and French, although the last one is not as widely spread. About 25% of Canadians use French as their home language (Statistics Canada, 2020), which is consistent with our findings from the PISA 2012 data (see Table 3).

As Table 1 shows, the population is polarized, with many provinces sharing a low percentage of the total. Most of Canada is concentrated in Ontario and Quebec, each one being the main representative of their respective official language: English and French. Quebec concentrates the majority of francophones along with New Brunswick, being it the only territory that has both languages as official.

According to the Section Sixteen of the Canadian Charter of Rights and Freedoms (Constitution Act, 1982), English and French are the two official languages of the State, meaning that both have equality of status in the Parliament, courts and Government of Canada. At the province level this is different, as usually only one is considered the official, although some of the provided public services are available in both of them.

Given all of this, all schooling across Canada is implemented in both languages, but in different proportions depending on the province. The majority of it is publicly funded (see Table 2). The educational system is divided into primary, secondary and post secondary education, although it varies on the territory, as it is under exclusive provincial jurisdiction (CICIC, 2020). For example, education is compulsory up to the age of 17 years all throughout the country, except for Ontario, Manitoba and New Brunswick, where it is 18 years.

Regarding the 2012 PISA results, Canada stands as one of the countries with the highest scores, being in the top 10 of the best performances in mathematics, reading and sciences (Brochu, Deussing, Houme & Chuy, 2013). Among all the territories in Canada, Quebec was placed in the top-performing participants globally in paper-based mathematics. In sciences, students in British Columbia and Alberta performed better than the country's average.

Table 1: *Population by mother tongue and geography, 2016*

Province	Population	Official Lang.	Speakers by Mother tongue (%)	
			English	French
Ontario	13,312,870	English	68.2	4.0
Quebec	8,066,555	French	8.1	78.0
British Columbia	4,598,415	English	70.0	1.4
Alberta	4,026,650	English	75.4	2.0
Manitoba	1,261,620	English	72.6	3.4
Saskatchewan	1,083,240	English	83.2	1.5
Nova Scotia	912,300	English	91.4	3.4
New Brunswick	736,285	Both	64.8	31.9
New F. and L.	515,680	English	97.1	0.5
Prince Edward Is.	141,020	English	91.1	3.6

Source: Statistics Canada.

2 Literature Review

During the last century, authors have studied the variables and factors affecting the school choice, although few has considered the coexistence of more than one language in their models. Bilingualism in Canada is key to understand the current situation and how it has affected its education system throughout the centuries (Vaillancourt, 2012). In Canada, language and religion have been always intrinsically related, as 70% of the French speakers in 1871 were Catholic. This fact led to a historical indirect protection of the language through religion, which started with the establishment of the Constitution Act of 1867, whose section 93, was introduced as a measure of protecting those Catholic Communities. This stated that the current and future school systems of the religious minority (which could be either protestant or catholic), existing in a province, should be protected from any potential harm from its provincial government; and gave authority to the Government to intervene to protect such minority.

2.1 The Influence of Official Minority Languages

The introduction of the Constitution Act led to situations such as the one in New Brunswick, which had a very large percentage of Catholics in 1871, and which nowadays share both languages as official ones. In Ontario, although having a very small linguistic minority, there exists an above average constitutional protection, as half of the population spoke French at the time. This protection was necessary, as there have been always a unwillingness of English-Canadians to collaborate in such purpose, giving the francophones in Canada an "added incentive to pursue nationalistic language policies, with the possibility of some consequences that English Canadians may not always appreciate", as Breton stated in his 1978 paper, "Nationalism and Language Policies".

As this system did not consider the language but the religious rights, in 1982 the "Canadian Charter of Rights and Freedom" was redacted, where the protection of minority languages educational rights got introduced into the legal system. Villancourt (2012), analyses the measurement of the costs and benefits of these official languages policies in

Canada, and explains that, although the coexistence of these two languages do not have a positive direct effect, it greatly increases the person welfare, as the public services are available in their mother's tongue.

Apart from this, Canadians may also have an increased utility from other types of benefits, such as the obvious cultural value increment of having more than one official languages, and indirect monetary effects. Christofides and Swidinsky (2008), found that men outside of Quebec (where the majority is francophone) with French language skills, tended to be disproportionally represented in higher paying occupations. Breton (1998), analysed census data from 1971-1991, concluding that there were significant returns to bilingual language skills in the estimated wage equations. Being a bilingual country have also its detriments, as the coexistence of the two languages can be very expensive regarding bureaucracy, as a consequence of the multilingualism (Fidrmuc, 2011; Pons-Ridler & Ridler, 1990).

The school system in Canada differs from other bilinguals' countries, such as Belgium or Switzerland, where the instruction language of the school depends on the region, and there is no real free choice. Similar to Basque Country, in Canada the parents can choose the language of education for their children. This concept is explored by Pons-Ridler & Ridler (1990), stating that the establishment of bilingualism in Canada, regarding the school choice, follows the "Personality Principle". According to that, the individual customer of Government services will be attended in the language of his choice. The authors also state that territoriality would be a cheaper alternative, in a similar way to Belgium and Switzerland, as it would save a lot of education expenditure.

Languages can be seen both as a means of communication and as part of a cultural identity as established by CASLT (2016). For instance, English is widely used in the business as a communication tool, and no cultural heritage is derived from its use in such environment. Specifically, CASTL (2016, pp. 62) states that, "Language, however, transmits not only meaning in a strict, terminological sense, but also commonly held moral

values, judgments of other social or political references". It is quite clear that the francophones in Canada have historically shared a common heritage, given how it was tied to the Catholic Community.

2.2 Previous Studies on School Choice

Manski and Wise (1983) are considered the first authors to model choice in the educational environment, but now the literature regarding school choice is abundant. For example, Nakhaie (2000) found that the parental education and occupation have significant consequences for their children's educational attainment. The study used log odd ratios to analyse two national representative samples of Canadians surveyed in 1985 and 1994. Likewise, Borgers et al (1999), use an extended logit model to analyse data from Veldhoven, in the Netherlands, which has two Catholic, one Protestant and two public schools. They found that the religion of the household, the size of the classroom, and the distance to the schools, were the most relevant variables. This is proved once again by Müller, Haase and Seidel (2012), by analyzing German schools data with a multivariate analysis to find that there exists an increasingly competitiveness among all schools, as a result from the possibility of free school choice. Their approach explicitly accounts for spacial substitution, as the authors considered that this patterns between school locations exist, setting a different approach from the previous literature, which is usually focused on racial mix, tuition fees, and travel-to-school distance.

Following the same steps, Hétu (1991) uses a multivariate analysis model to analyse gathered data from 1,894 French speaking students in Quebec, at grade 5, 10 and 12, where 12.7% assisted private schools, to gather different conclusions, such as: "[...] parents' assessment of the 'quality' of teaching in public and private institutions probably corresponds to a particular teaching style, moral values and cultural preferences rather than merely to school performance measured by grades and failures." (Hétu, 1991, pp. 496) This clearly shows that the culture is a key factor for parents when considering different school types.

More recently, Freynet and Clément (2015) also used a multivariate analysis of variance to assess the effect of bilingualism over the identity, proving that bilinguals are significantly distinct from predominantly monolingual participants on most factors related to the cultural identity. They used data collected by Statistics Canada in 2006, through computer assisted telephone interviews, as part of the Survey on the Vitality of Official Language Minorities, conducting an analysis of variance in order to determine the effects of the different variables.

Wyckoff, Lankford and Lee (1995), used 1,985 students level data from the Current Population Survey to estimate a random utility model of school. The empirical results showed that the most important variables were the racial composition of the school, the crime rate, the religious orientation, as well as socio-economic characteristics of the student and the family. Similar results were obtained by Prieto, Aguero-Valverde, Zarrate-Cárdenas and Van Maarseveen, (2018), who used data from a large school district in Florida, to design discrete choice models to analyse the impact of different factors regarding specialty school choice. They replicated the previous Wyckoff et al (1995) findings concluding that: "Having a school with a large proportion of students of the applicant's own race is associated with an alternative being preferred by the applicant" (Prieto, Aguero-Valverde, Zarrate-Cárdenas and Van Maarseveen, 2018, pp. 22. They also found that the students tend to self-segregate, even when they have no explicit desire to do so. Similar results were found by Bifulco and Ladd (2007), when using panel data to examine the effects of charter schools in North Carolina on racial segregation and white-black score gaps. Moreover, they discovered segregation and achievement gaps, when discriminating by the parent's socioeconomic status.

Immigration is also a relevant factor as shown, e.g. by Farre et al (2018), who analysed Spanish data with OLS and tobit models. He found that the immigration caused a native flight from the public schools (even more acute on the more educated households), and a higher student-to-teacher ratio in public schools, widening the gap between the educational investments of rich and poorer households.

There have been research from other interesting points of view, such as the paper from Maddaus (1990), which delves into the parental choice of school, discovering that they have a more holistic view than what policy makers believe (moral, social, emotional, cognitive...) and that wealth affects how informed they are when choosing their school. Calsamiglia, Haeringer and Klijn (2010), introduced constraints in the choice in order to approximate better the reality in the form of neighbourhoods, limiting the eligible schools. Nevertheless, the introduction of bilingualism as a key factor in the school choice modelling is a fairly recent event.

We can conclude by saying that, the findings of previous literature show unremarkable evidence of social class segregation in the school choice: wealthier more educated families will send their children to private school, and the more modest ones to a public one. Moreover, wealthier parents will be able to send their kids to they preferred school - even if it is not private - as they will be able to afford indirect expenses. A recent paper by Vega-Bayo, Sanko and Mariel (2019), reaches to the same conclusion when studying the social class segregation in upper-secondary school choice in Japan. They make use of a multinomial logit to find that the most influential family characteristics in the upper-secondary school choice in Japan are the number of children, the cultural goods - acting as a proxy of family wealth - and the parent's education level.

2.3 A different approach: bilingualism

Mariel and Vega-Bayo (2015), analysed PISA data from the Basque Country, which has several types of schools depending on their financing and their instructional language, using a multinomial logit model. They found that, otherwise very important socio-economic variables, as the previous literature proves, are overshadowed by the effect of the bilingualism, which mitigates class stratification. This outcome is what we expected to found in the Canadian dataset analysis, where sociodemographic variables as the educational level and the family wealth are relegated to the background.

In a more recent study, Mariel and Vega-Bayo (2018a) performed similar analysis for the rest of regions of Spain, finding that overall, the most important variables for the school choice are the wealth and the socio-economic status of the families, results previously proven in the literature. In that line, Mariel and Vega-Bayo (2018b) carried out a specifically designed discrete choice experiment in the Basque country to analyse which school characteristics the parents prefer in the Basque Country when choosing a primary school for their children. Their results confirms previous findings regarding the Basque Country showing that the social segregation in the school choice is diluted by the bilingualism.

3 The Data Analysis

This section will be devoted to the analysis of the data we extracted from the PISA 2012 database. For this purpose, we will go through some basic descriptive statistics of the database, along with the division of the provinces regarding private schooling.

3.1 Description of the Database

The data used in this project have been extracted from the 2012 PISA study (OECD, 2012). This programme consists on an extensive analysis of the acquired skills of high school students regarding key subjects, such as Mathematics or Science. A total of 16,557 individuals form the sample, as we center our analysis on Canada. As can be seen, the vast majority of the understudies were in a government funded school (see Table 2).

Table 2: *Number of Surveyed Students*

Type of School	Students	Percentage
Public English	11,178	67%
Private English	610	7%
Public French	4,324	26%
Private French	445	3%
	16,557	100

Source: OECD PISA 2012, Canada.

Given that in Canada there are anglophone and francophone school systems, a larger sample was needed in order to get reliable results (Brochu, Deussing, Houme & Chuy, 2013). For most provinces, samples were recollected for each framework. Table 1 shows the percentage of speakers for each language and as shown, in most of the provinces English is the official language; with the exception of Quebec, where the francophones compounds the majority, and New Brunswick. By taking into account this information, we can comprehend why the sampling in Canada implied more exertion.

In the PISA 2012 investigation, the *STRATUM* variable was the instrument which we used to distinguish between instructional languages, as it explicitly segregates by language, territory and school size. Disregarding this, some of the provinces did not follow this distribution and thus, were excluded from our models: Prince Edward Island and Newfoundland & Labrador.

Table 3: *Summary statistics of the explanatory variables*

Variable				
Dummy variables	Proportion			
Home language not English	0.32			
Mother not working	0.35			
Two-parent family	0.81			
Siblings	0.81			
Grandparents living with the family	0.07			
Continuous variables	Mean	S.D.	Min.	Max.
Cultural possessions	-0.2	0.9	-1.5	1.3
Educ. level mother	4.9	1.3	0.0	9.0
Educ. level father	4.8	1.5	0.0	9.0
Family wealth	0.5	1.0	-5.5	2.9
Highest parental occupational status	55.3	20.7	11.0	88.9

Source: OECD PISA 2012.

Table 3 includes the summary statistics for all of the independent variables considered in our model. As a large portion of those are dummy or binary variables, whose outcome is either 1 or 0, only the mean is shown (the proportion). The results shows that a 32% of the households do not speak in English at home. This variable will be vital to evaluate the impact of bilingualism over parental choice, for each school system. Moreover, most of the households have more than one child, sharing a very similar proportion with the biparental families. Finally, partaking the household with grandparents is a very rare event.

All of the continuous variables used as explanatory in our model, that are presented in Table 3, are index values constructed by PISA 2012. For instance, *Cultural Possessions* showcases the cultural level of the families, by taking into account all of the artistic related goods in the households. This acts as a proxy for the socio-economic status, as variables like wealth, can be insufficient for this purpose. As seen in Table 3, the values of *Wealth* range from -5.48 to 2.90, measuring the properties of the family: general home possessions plus country-specific wealth items; such as a link to the internet, cars, computers, or in the case of Canada, an ipod.

Analogously, the educational level of the progenitors is defined by the *International Standard Classification of Education* (ISCED) of 1997, where the lowest value, 0, indicates no education at all, and the maximum value, 6, specifies second-stage tertiary education: a master's degree or a PhD. Finally, the last variable measures the highest occupational status of parents, according to the *International Socio-Economic Index of occupational status* (ISEI).

3.2 Division of the provinces

When we analyzed the data, we found that half of the regions showed few to none private schools (see Table 5). Therefore, we split our study into two subsets to perform two different analysis. In the first subset we include those provinces that had enough observations for all four outcomes: French and English, public and private schools. Thus, the discrete choice methodology behind this analysis is a multinomial logit model. The codification for each outcome, as computed in the model, is respectively: 1 for the *Public English*, 2 for the *Private English*, 3 for the *Public French*, and 4 for the *Private French*.

In the second group, we discarded individuals that assisted to private schools, as the size of the data was insufficient to perform the analysis. Hence, we created a binary logit model with only two outcomes: *Public English* or *Public French*. In each subset, we can find a province expected to have a higher influence on the language component, given its francophone population: Quebec in the first, and New Brunswick in the second. For this subset, the dependent variable will take value 1 if the instructional language of the school is French, and 0 if English.

Table 4: *Group I: Distribution of public, private, English and French school.*

	Quebec		Manitoba		Saskatchewan		British Columbia	
Type	n	%	n	%	n	%	n	%
Public English	1,165	34.1	1,481	80.5	1,588	90.7	1,409	86.4
Private English	293	8.6	94	5.1	74	4.2	73	4.5
Public French	1,713	50.2	204	11.1	32	1.8	92	5.6
Private French	243	7.1	60	3.3	56	3.2	57	3.5
Total Private	536	15.7	154	8.4	130	7.4	130	8.0

Source: OECD PISA 2012, Canada.

Table 5: *Group II: Distribution of public, private, English and French school.*

	Nova Scotia		Alberta		New Brunswick		Ontario	
Type	n	%	n	%	n	%	n	%
Public English	1,064	86.6	1,582	89.0	820	59.6	2,069	63.4
Private English	-	-	-	-	28	1.7	48	1.5
Public French	165	13.4	166	9.3	805	48.7	1,147	35.1
Private French	-	-	29	1.6	-	-	-	-
Total Private	-	-	29	1.6	28	1.7	48	1.5

Source: OECD PISA 2012, Canada.

4 Methodology

In this section, we will set the theoretical foundation and the description of the discrete choice models we utilized in the analysis of the database: the multinomial and binary logit.

4.1 Theoretical framework

As our dependent variable is a categorical one with 2 or 4 outcomes - depending on the model- we apply a discrete choice method to analyse the likelihood of an individual picking a specific alternative. Hence, we model the parent's school choice by a multinomial logit model (MNLM) in provinces where four type of schools are available (English Private, English Public, French Private, French Public) and binary logit model in provinces with only two alternatives (English Public, French Public).

For the first subset, we have created a binary logit, which only has two outcomes, *Public English* and *Public French*. The dependent variable will take value 0 if the instructional language of the school is English, or 1 if it is French. As well known, the equation below is the formula for the predicted probability of outcome 1 (Long, 1997).

$$Pr(y = 1|x_i) = F(x_i'\beta) = \frac{exp(x_i'\beta)}{1 + exp(x_i'\beta)}, \tag{1}$$

What we are modelling it's the probability of choosing French as the instructional language given the sociodemographic characteristics of the student. In our study, the reference level is outcome 0, that is, the student assists to a Public English school. In the formula, β stands for the parameters to be estimated, and x_i is the vector containing the sociodemographic characteristics of the student (see Table 3).

For the second subset, we have estimated a multinomial logit model, which has four outcomes instead: *Public English*, *Private English*, *Public French* and *Private French*. These are computed in the model as $y = 1, 2, 3, 4$, respectively. Being y the dependent variable with $J = 1, 2..., 4$ outcomes, and $Pr(y = m|x)$ the probability of observing m as the outcome given x_i, with this multinomial logit we are estimating the probabilities of choosing each school type. Hence, there will be 4 estimations for each observation. Equations below includes the formula for the predicted probability of outcome m. We need these two formulae in order to identify the model, so the joint sum of all probabilities equals one. The easiest way of doing this, is by imposing the restriction of setting all the $\beta's$ from one category to 0, in this case when the dependent variable is equal to 1. The reference level is, once more, a student that assists to a *Public English* school.

$$Pr(y_i = 1|X_i) = \frac{1}{1 + \sum_{j=2}^{J} exp(X_i'\beta_j)}, \tag{2}$$

$$Pr(y_i = m|X_i) = \frac{exp(X_i'\beta_m)}{1 + \sum_{j=2}^{J} exp(X_i'\beta_j)} \text{ for m} < 1. \tag{3}$$

This equations will be the basis for the Maximum Likelihood estimator, which will be the method used for the estimation of the parameters, for both models. Moreover, is not wrong to consider the multinomial logit model as "a linked set of binary logits" (Long, 1997, pp. 151).

4.2 Interpretation of the model: discrete changes in probabilities

As Long (Long, 1997, pp. 164) states, there are many parameters in a multinomial logit. Although the statistical significance is important, the magnitudes and the direction of the effects are ignored. In his book, he describes several methods of interpreting the results of the multinomial model, such as *predicted probabilities* and *partial change*. In this research, we have used *discrete change* to measure the changes in the probabilities. This works for continuous and dummy variables.

The predicted probability of the dependent variable y, being equal to the outcome m given the vector of covariates x_i is:

$$Pr(y = m|x_i) = \frac{exp(x_i'\beta_m)}{\sum_{i=1}^{J} exp(x_i'\beta_j)} \tag{4}$$

By changing the x_k parameter from x_S to x_E, *ceteris paribus*, the change in the predicted probability is:

$$\frac{\Delta Pr(y = m|x)}{\Delta x_k} = Pr(y = m|x, x_k = x_E) - Pr(y = m|x, x_k = x_S). \tag{5}$$

The magnitude of the change in probabilities depends on the amount of change in x_k, the starting value of x_k and the other variables (Long, 1997). Most of the time, the other x_{-k} variables are held at their mean values, with all the dummies at either 1 or 0. How much should x_k vary? If it is a dummy, the choice is pretty straightforward, as it is a discrete change from 0 to 1, or vice versa. If on the contrary, the chosen variable is continuous, the amount of change depends on your analysis and on the nature of it. In section 5.3 this will be further discussed.

Usually the best available tool for this type of analysis is a *discrete change plot*, as it quickly summarizes all the information. In this type of graph, the horizontal axis represents the magnitude of the positive or negative change in the probability for each outcome. These are represented by letters. The vertical axis includes all the variables of the model (see fig. 1 and fig. 2). This analysis tool is suitable for both binary and multinomial logits.

5 Empirical Results

This section contains all of the obtained results from our analysis, from the definition of our model to the interpretation of our estimations.

5.1 Our models

As explained in section section 3.2, we have split the database into two, as there were two regions clearly defined by its private schooling percentage. Each of the subconsequent models have different variables, because we have distinguished the provinces inside them, as well as established differential effects for the key variables: family wealth and the spoken language in the household. For the estimation of the model, we adopt the *R* programming language, more specifically, the *VGAM* and *NNET* packages. Moreover, as we cannot interpret the coefficients directly, due to the non-linear nature of the model, we interpret our results using discrete changes in probabilities, as explained in section section 4.2. The results are shown in the appendix, in Table 8 and Table 9. Posteriorly, both formulas are shown.

Hence, for the first group of provinces, we have defined a multinomial logit model as the tool to analyse the variables that drives Canadian parents to choose a specific type of school, in a bilingual environment. Our dependent variable, *type of school*, has four categories: public English, private English, public French and private French. As previously said, at Table 3 you will find the summary statistics for the explanatory variables, divided in continuous or dummy categories. This variables were based on the previous research done by Mariel and Vega-Bayo (2015, 2018a, 2019), regarding the subject. For the second group of variables we make use of a binary logit, where we estimate the probability of selecting a French speaking school compared to a English one.

$$Pr(SchoolType = m|x_i) = F(\beta_1 + \beta_2 Quebec + \beta_3 B.Columbia + ...) \quad (6)$$

$$= \frac{exp(x_i'\beta_m)}{\sum_{i=1}^{J} exp(x_i'\beta_j)}, \quad (7)$$

where $\beta_1 = 0$, as the restriction is needed for the proper identification of the model (see section 4.1), and x_i is the given vector of the covariates, and m is a specific outcome among the four options: Public English School, Private English School, Public and Private French Schools. For the second group of province, a binary logit model is estimated.

$$Pr(SchoolType = French|x_i) = F(\beta_1 + \beta_2 Alberta + \beta_3 NewBrunswick + ...) \quad (8)$$

$$= \frac{exp(x_i'\beta_m)}{1 + exp(x_i'\beta_m)}, \quad (9)$$

5.2 Relevance of the variables: Wald test

In order to test the significance of all the parameters associated with the outcomes of both models, we have chosen the Wald test to analyze the significance of the coefficients. The null hypothesis of this tests, is that the explanatory variable in question does not affect the explained variable, which involves three parameters (or only one in the second model). If at least one of the parameters is significant, the null hypothesis will be rejected. Table 6 and Table 7 show the results from this analysis. Most of them are significant at the 1% level, although three variables in the binary logit are not relevant at all, but we are keeping them as they are relevant to the first model, for comparison purposes: *Siblings*, *the educational level of the father* and *two-parent family*.

Table 6: *Wald tests for the significance of the explanatory variables (I).*

Variable	χ^2 statistic	*p-value*
Mother not working	42.3	$<$0.01 ***
Two-parent family	13.3	$<$0.01 ***
Siblings	31.1	$<$0.01 ***
Grandparents living with the family	15.3	$<$0.01 ***
Cultural possessions	128.9	$<$0.01 ***
Educ. level mother	21.7	$<$0.01 ***
Educ. level father	17.3	$<$0.01 ***
Family Wealth	108.4	$<$0.01 ***
Highest parental occup. status	122.7	$<$0.01 ***
Home language not English	1242.2	$<$0.01 ***
Quebec	284.3	$<$0.01 ***
Sasketchwan	63.2	$<$0.01 ***
British Columbia	58.7	$<$0.01 ***
***, **, * denotes significance at the 1%, 5% and 10% level respectively.		

Table 7: *Wald tests for the significance of the explanatory variables (II).*

Variable	χ^2 statistic	*p-value*
Mother not working	26.8	$<$0.01 ***
Two-parent family	13.3	$<$0.952
Siblings	0.1	0.335
Grandparents living with the family	35.9	$<$0.01 ***
Cultural possessions	116.5	$<$0.01 ***
Educ. level mother	9.8	$<$0.01 ***
Educ. level father	1.9	0.167
Family Wealth	7.3	$<$0.01 ***
Highest parental occupp. status	36.8	$<$0.01 ***
Home language not English	1247.3	$<$0.01 ***
Ontario	66.8	$<$0.01 ***
New Brunswick	114.4	$<$0.01 ***
Alberta	41.4	$<$0.01 ***
***, **, * denotes significance at the 1%, 5% and 10% level respectively.		

5.3 Discrete changes in the Probabilities

This last section of the empirical results includes the overview of the discrete changes in the probabilities and thus, is key for the interpretation of our model. Results include, for each group of provinces (there are two, as seen in section 3.2), an analysis for each explanatory variable. The theory behind this methodology is explained in section 4.2.

As previously mentioned, there are both dummy and continuous variables, and each type follows a distinct technique when studying its influence. In the first case, the change is intuitive: we will switch the initial value of the parameter from 0 to 1 or vicersa, depending on the benchmark. For the latter, we will add and subtract the standard deviation to the mean. This is not arbitrary, the logic behind this is that, presuming a normal distribution, the majority of the values will be included in this range.

First, results for the Group I of provinces will be commented, that is, those that showcased enough data for private schooling analysis, and then the second group, where we only discriminated the schools by language. The outcomes will be illustrated in fig. 1 and fig. 2. The tables containing the coefficients and the std. errors for each group can be found in the appendices.

5.3.1 First model: *Group I of provinces*

Before delving directly into the matter, we should recapitulate and take a look at this first group. It includes Quebec, Manitoba, Saskatchewan and British Columbia. Quebec is the second biggest province, both financial and population wise (Statistics Canada, 2016). It is also the principal francophone territory. British Columbia and Manitoba are not as big as the latter, but they are not small either, being the third and the fifth provinces by populace (see Table 1).

The setting of the benchmark should not be an arbitrary decision, as it will surely affect the changes in the probabilities. Having analyzed the magnitude of the explanatory variables, we have decided the following:

(a) The student lives in Manitoba, as Quebec, Saskatchewan and British Columbia had a bigger differential effects when introduced as dummies (therefore these three dummies are set to zero).

(b) The home language is English, as it is the most common outcome.

(c) The continuous variables are set to the mean, these includes the family wealth, the occupational status, and the educational level for both mothers and fathers.

(d) Finally for the the other dummy variables we have chosen the most common outcome, with siblings and with no grandparents living in the family unit.

Figure 1 showcases the changes in probability for each variable. In the upper right corner you can find the legend for the types of schools. The *R* language did not have an implemented library for this purpose (unlike Stata), so everything in the process is self coded.

Starting with the continuous variables, we can appreciate that they are not very prominent in general. The *cultural Possessions*, *occupational status* and the *educational level* of both parents show little to none influence over the probability changes. In spite of this, the *family wealth* parameter gains weight when we take in account the differential effects of the provinces. In both Saskatchewan and British Columbia the gain is noticeable, but not as huge as in Quebec. We can appreciate that the probability of choosing a public school (*1* and *3*) decreases, implying that for richer families, private schools are preferred. For Saskatchewan, public French schools have a higher probability of being chosen, this could come because of the interest of richer households of their children speaking more

than one language. Christofides and Swidinsky (2008) found that men outside Quebec that spoke French, were disproportionally represented in higher paying jobs. *Cultural Possessions* have a similar effect with the family wealth, albeit smaller, with private schools situated at the right side.

Having analyzed the dummy variables we can distinguish several subgroups: those related to bilingualism, to the territory, and lastly the socioeconomic ones. Starting with the provinces, the only significant changes in the probabilities come from Quebec and British Columbia, with the first one being noticeable bigger. As the benchmark family speaks English, the probability of choosing a French public school is reduced. Also, as Quebec has twice the number of students from private schools, the discrete change is situated at the right side. Regarding the socioeconomic dummy variables, we cannot excerpt much, as the discrete changes are less than 5%.

The most relevant discrete changes are those related to bilingualism. The variable *Home Language not English* is also combined with the provinces to show the differential effects. In the case of Quebec, the probability of assisting to public French schools is increased by a large margin, a logical consequence of its francophone population, while the chance of going to a public English school is greatly diminished.

5.3.2 Second model: *Group II of provinces*

As previously done, lets take a look into the composing provinces of the second group: Nova Scotia, Alberta, New Brunswick and Ontario. Ontario is the most populated territory in Canada, holding approximately 40% of the state. Alberta is the fourth largest province, and the others two have a low share of the total population (see Table 1). They all have English as the official language, with the exception of New Brunswick, which have both. Lastly, they all have in common a low number of private schools, which made us split the provinces. Because of this, we will be using the binary logit model, as explained in section 4.1.

The setting of the benchmark is not an arbitrary decision, as it will surely affect the changes in the probabilities. Having analyzed the magnitude of the explanatory variables, we have decided the following:

(a) The student lives in Nova Scotia, as Quebec, Saskatchewan and British Columbia had a bigger differential effects when introduced as dummies (therefore these three dummies are set to zero).

(b) The home language is English, as it is the most common outcome.

(c) The continuous variables are set to the mean, these includes the family wealth, the occupational status, and the educational level for both mothers and fathers.

(d) Finally for the the other dummy variables we have chosen the most common outcome, with siblings and with no grandparents living in the family unit.

(e) The reference level of the binary logit is *Public English*.

Figure 2 showcases the changes in probability for each variable. In the upper right corner you can find the legend for the types of schools. The *R* language did not have an implemented library for this purpose (unlike Stata), so everything in the process is self coded.

As previously stated in section section 3.2, for this group of provinces we only study the election of public schools. Due to this, the logit model is the right tool for our analysis. Regarding the continuous variables, very similar results rise from this binomial model, as their influence on the outcome is rather lackluster - for most of them. Among these, the most prominent variables is *Family Wealth*. Its differential effects captures the diverse amalgam of the North American country. As a matter of fact, Ontario reveals the humongous importance of the family wealth when compared to the others provinces, across both models. In spite of the low percentage of French speakers my mother tongue, it has a re-

markable share of public French schools - about 35% of the questioned students assisted to one - as presented in Table 3, much higher than the rest of the official English territories.

Moreover, this high incidence is also captured by its dummy, as a consequence of that high number of French schools, in spite of the low number of households speaking it. The other provinces have little to none weight on the outcome. Moving to the other socioeconomic dummy variables, we find that most of them excerpts no influence whatsoever, except maybe *Cultural possessions* and *Grandparents*.

Analogically to the obtained results of the first model, bilingualism is the main determinant of the school choice outcome. The differential effects vary considerably, even more than in the previous model. By considering this territorial discrimination, the probability of assisting a Public French school increases, approximately in 0.8, if the family is from New Brunswick and does not speak English at home. This confirms the high influence of having both languages as official.

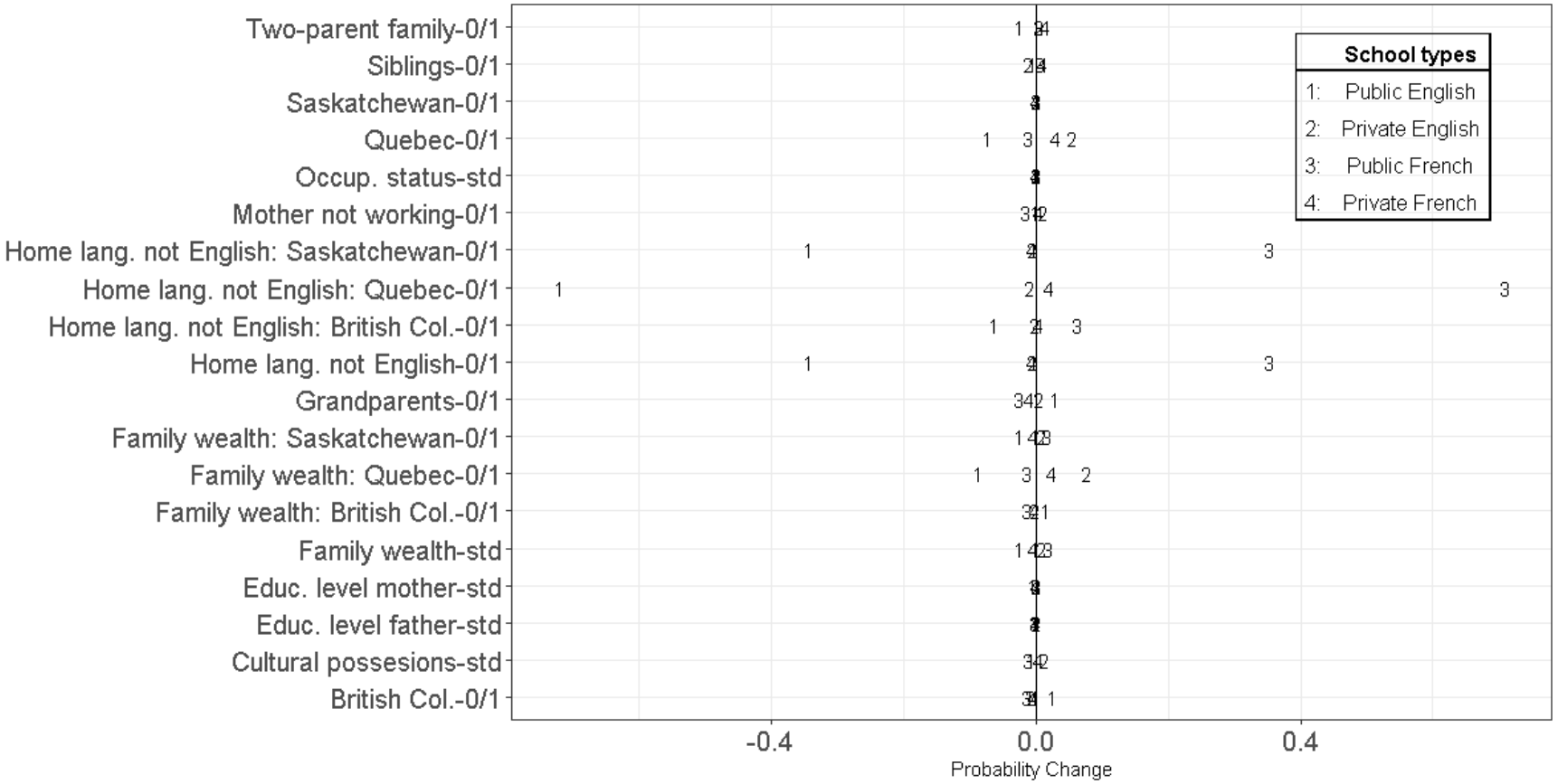

Figure 1: *Change in probability with respect to the benchmark family, Group I*

The vertical line represents the *benchmark family:* from Manitoba, Home language English, mean Wealth, mean Occup. Status, mean Cultural Possessions, mean Educational Level for the father and mother (not working), students has Siblings and no Grandparents.

Source: self-made with R.

Universidad del País Vasco
Euskal Herriko Unibertsitatea

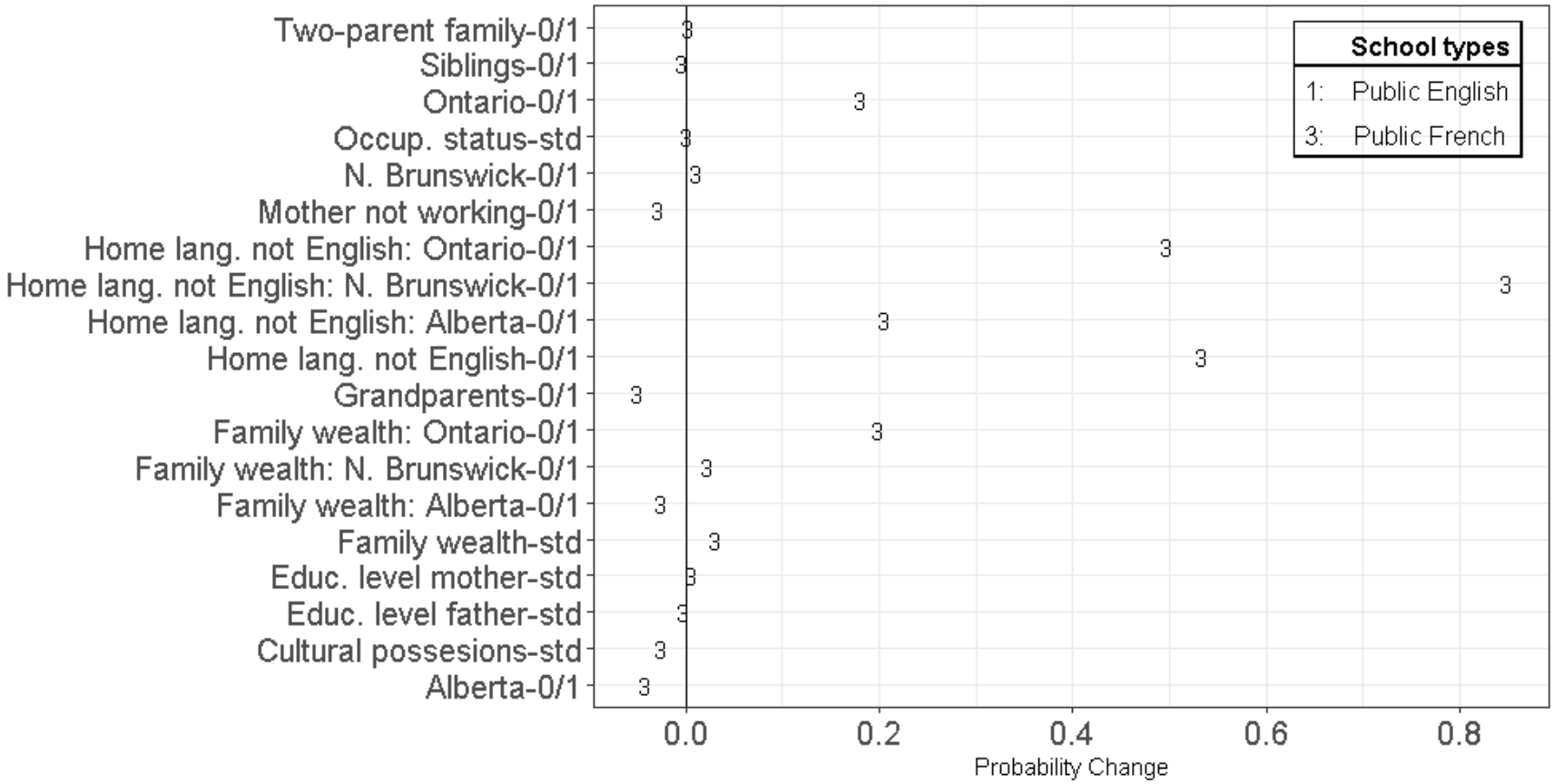

Figure 2: *Change in probability with respect to the benchmark family, Group II*

The reference level of the binary logit is *Public English.* The vertical line represents the *benchmark family:* from Nova Scotia, Home language English, mean Wealth, mean Occup. Status, mean Cultural Possessions, mean Educational Level for the father and mother (not working), students has Siblings and no Grandparents.

Source: self-made with R.

Universidad del País Vasco Euskal Herriko Unibertsitatea

6 Concluding Remarks & Further Research

This thesis analyses the social class segregation in upper-secondary school choice in Canada. This type of research had been done extensively before (Manski and Wise, 1983; Hétu, 1991; Wyckoff et al, 1995, Calsamiglia et al, 2010, among many others), but the introduction of bilingualism on the model is fairly recent. (Mariel et al Spanish study, 2015). The main findings of this previous literature are that there is an unremarkable evidence of social class segregation in the school choice, as wealthier households will select private and the best among the public schools, as they will be able to suffrage the indirect expenses (Bifulco and Ladd, 2007). What we wanted to study in this thesis, is if the bilingualism affects the social class segregation in the same way as in the Mariel et al Spanish study (2015).

Hence, in order to answer this question, we selected the PISA 2012 questionnaires, as it is currently one of the most complete databases regarding students assessment. It contained enough information for us to be able to discriminate the observations by instructional language, which is key due to the nature of our analysis. The surveyed sample is representative of the population, as the numbers match with the results from the Statistics Canada census (2016), with regards to the population by mother tongue and geography. In spite of this, we were not able to study two provinces: New Foundland & Labrador, and Prince Edward Isle, as they did not stratified the schools by language. This is not a problem, as they only contain less than 2% of the total population of Canada.

For our analysis, we constructed our discrete choice analysis with binary and multinomial logit models. The two methodologies were necessary as one half of the provinces show little to none private schooling, and the results were not sufficient. The first group of provinces include Quebec, Manitoba, Saskatchewan and British Columbia; while the second contains Nova Scotia, Alberta, New Brunswick and Ontario. Among these, we have one representative of French, the minority language in Canada, for each group: Quebec and New Brunswick (see table Table 1). The latter has both languages as official.

Overall, the estimation of the parameters from both models are robust. The Wald test confirms that all variables are relevant across both models. The probability of choosing schools where French is the instructional language is affected the most by those variables related to the home language. This proves that bilingualism is the main driver of school choice. Among the provinces, Quebec and New Brunswick uncover the greater influence of French when compared to the others. Moreover, the second most important group of parameters are those related to family wealth. Once again Quebec showcases a bigger differential effects, as well as Ontario, which is logical given that they are among the richest territories in Canada (Statistics Canada, 2016). Having analyzed all of the variables, some of them stand out for their lack of relevance, such as *siblings* and the *educational level* of both parents, which are relevant in other papers, such as in the Mariel et al study of Japan (2019).

In conclusion, we have gathered compelling evidence of bilingualism diluting the socioeconomic segregation previously found in the school choice. It would be interesting to analyze other bilingual or multilingual countries. However, we need to take into account that not many countries boast from the unremarkable situation of Canada or the Basque Country. In order to properly understand the bilingualism effect on the school choice, the instructional language must not be restricted by territory (e.g. Belgium, where the language depends on the area in which you study).

There are 55 officially bilingual countries in the world (University of Ottawa, 2020), so there is plenty of potential research ahead. Regarding Canada, it would also be alluring to use other databases to test our findings. *Statistics Canada*, is the official institute of statistics of the state, and it contains enough studies to carry a similar analysis to the one we have conducted, such as the *Classification of Instructional Programs* (Statistics Canada, 2016). Studying school choice from another point of view could be interesting: as a compromise between fulfilling the preferences of the households (people will tend to segregate on their own), or making an effort to eradicate the existing segregation on a social, racial, or cultural level.

References

[1] Bifulco, R., & Ladd, H. F. (2006). School choice, racial segregation, and test-score gaps: Evidence from North Carolina's charter school program*. *Journal of Policy Analysis and Management, 26*(1), 31-56.

[2] Borgers, A., Oppewal, H., Ponjé, M., & Timmermans, H. (1999). Assessing the Impact of School Marketing: Conjoint Choice Experiments Incorporating Availability and Substitution Effects. Environment and Planning A: *Economy and Space, 31*(11), 1949-1964.

[3] Breton, A. (1978). Nationalism and Language Policies. *The Canadian Journal of Economics, 11*(4), 656.

[4] Breton, A. (1998). *Economic approaches to language and to bilingualism*. Ottawa: Canadian Heritage.

[5] Brochu, P., Deussing, M., Houme, K. & Chuy, M. (2013). *Measuring up: Canadian results of the OECD PISA study: The performance of Canada's youth in science, reading and mathematics: 2012 first results for Canadians aged 15*. Council of Ministers of Education, Canada.

[6] Calsamiglia, C., Haeringer, G., & Klijn, F. (2010). Constrained School Choice: An Experimental Study. *American Economic Review, 100*(4), 1860-1874.

[7] Christofides, L. N., & Swidinsky, R. (2010). The Economic Returns to the Knowledge and Use of a Second Official Language: English in Quebec and French in the Rest-of-Canada. *Canadian Public Policy, 36*(2), 137-158.

[8] CICIC, 2020. *Ministries/departments responsible for education in Canada.* Retrieved from the CICIC webpage.

[9] Constitution Act, 1982. Retrieved from the Government of Canada webpage.

[10] Economic Advantages of Bilingualism. Literature Review (2016). *Canadian Association of Second Language Teachers*.

[11] Farre, L., Ortega, F., & Tanaka, R. (2018). Immigration and the public–private school choice. *Labour Economics, 51*, 184-201.

[12] Fidrmuc, J. (2011). The Economics of Multilingualism in the EU. *Research Handbook on the Economics of European Union Law*.

[13] Freynet, N., & Clément, R. (2015). Bilingualism in minority settings in Canada: Integration or assimilation? *International Journal of Intercultural Relations, 46*, 55-72.

[14] Ganzeboom, H. B., Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. Social Science Research, 21(1), 1-56.

[15] Hétu, C. (1991). Educational reform and tradition: The case of public and private high schools in Quebec. *International Journal of Educational Research, 15*(5), 483-497.

[16] Lankford, R., Lee, E., & Wyckoff, J. (1995). An Analysis of Elementary and Secondary School Choice. *Journal of Urban Economics, 38*(2), 236-251.

[17] Long, J. S. (1997). *Regression Models for Categorical and Limited.* Dependent Variables, SAGE Publications, Thousand Oaks.

[18] Maddaus, J. (1990). Parental Choice of School: What Parents Think and Do. *Review of Research in Education, 16, 267*.

[19] Manski, C. F., & Wise, D. A. (1983). *College choice in America*. Cambridge, Mass. u.a.: Harvard Univ. Pr.

[20] Müller, S., Haase, K., & Seidel, F. (2012). Exposing Unobserved Spatial Similarity: Evidence from German School Choice Data. *Geographical Analysis, 44*(1), 65-86.

[21] Nakhaie, M. (2000). Social origins and educational attainment in Canada: 1985 and 1994. *Review of Radical Political Economics, 32*(4), 577-609.

[22] OECD. (2012). *Programme for International Student Assessment* (PISA).

[23] Pons-Ridler, S., & Ridler, N. B. (1989). The territorial concept of official bilingualism; A cheaper alternative for Canada? *Language Sciences, 11*(2), 147-158.

[24] Prieto, L. M., Aguero-Valverde, J., Zarrate-Cardenas, G., & Maarseveen, M. V. (2018). Parental preferences in the choice for a specialty school. *Journal of School Choice, 13*(2), 198-227.

[25] Ripley, B. & Venables, W. (2020). *Feed-Forward Neural Networks and Multinomial Log-Linear Models*. R package, version 7.3-14.

[26] Statistics Canada (2020). Table 15-10-0003-01. *Population by mother tongue and geography, 1951 to 2016.* Retrieved from the Statistics Canada webpage.

[27] Statistics Canada (2016). Aboriginal Population Profile. *2016 Census*. Retrieved from the Statistics Canada webpage.

[28] Statistics Canada (2016). Language Highlight Table. *2016 Census*. Retrieved from the Statistics Canada webpage.

[29] Statistics Canada (2016). Table: 36-10-0487-01. *Gross domestic product (GDP) at basic prices, by sector and industry, provincial and territorial* . Retrieved from the Statistics Canada webpage.

[30] Statistics Canada (2016). *Classification of Instructional Programs (CIP) Canada 2016*. Retrieved from the Statistics Canada webpage.

[31] University of Ottawa (2020). *Compendium of Language Management in Canada (CLMC)*. Retrieved from the University of Ottawa webpage.

[32] Vaillancourt, F. (2012). *Official language policies of the Canadian provinces: Costs and benefits in 2006*. Vancouver, BC: Fraser Institute.

[33] Vega-Bayo, A., & Mariel, P. (2015). School choice in the Basque Country: Public, government-dependent and private schools with different languages of instruction. *International Journal of Educational Research, 74*, 13-25.

[34] Vega-Bayo, A., & Mariel, P. (2018). School Choice across Different Regions of Spain. *Revista Hacienda Pública Española, 227*(4), 11-36.

[35] Vega-Bayo, A., & Mariel, P. (2018). A Discrete Choice Experiment Application to School Choice. *Revista Hacienda Pública Española, 230*(3), 41-62.

[36] Vega-Bayo, A., Sanko, N., & Mariel, P. (2019). Social class segregation in upper-secondary school choice in Japan. Universidad del País Vasco, Departamento de Economía Aplicada III.

[37] Yee, T. & Moler, C. (2020). *Vector Generalized Linear and Additive Models*. R package, version 1.1-3.

Appendix I

Table 8: *Multinomial logit: first model estimation*

	Dependent variable:		
	Private English	Public French	Private French
	(1)	(2)	(3)
Mother not working	0.439***	−0.292***	0.226**
	(0.100)	(0.083)	(0.109)
Two-parent family	0.268*	0.108	0.617***
	(0.155)	(0.107)	(0.179)
Siblings	−0.471***	0.133	0.738***
	(0.135)	(0.108)	(0.190)
Grandparents	0.187	−0.560***	−0.589**
	(0.207)	(0.170)	(0.278)
Cultural Possessions	0.483***	−0.217***	0.229***
	(0.057)	(0.043)	(0.059)
Mother Educ. level	0.219***	0.035	0.127**
	(0.052)	(0.033)	(0.051)
Father Educ. level	0.149***	−0.029	0.055
	(0.042)	(0.027)	(0.042)
Family Wealth	0.522***	0.330***	−0.104
	(0.113)	(0.087)	(0.148)
Occup. Status	0.031***	0.010***	0.016***
	(0.003)	(0.002)	(0.003)
Home Language not English	0.231	2.413***	0.203
	(0.338)	(0.175)	(0.383)
Quebec	1.453***	0.266	1.142***
	(0.204)	(0.201)	(0.208)
Sasketchwan	−0.223	−1.678***	−0.099
	(0.267)	(0.331)	(0.263)
B. Columbia	−0.435	0.109	−0.439
	(0.267)	(0.202)	(0.278)
Home Language not English: Quebec	−0.668*	1.886***	1.093***
	(0.377)	(0.239)	(0.411)
Home Language not English: Saskatchewan	0.323	0.145	−0.628
	(0.575)	(0.432)	(0.830)
Home Language not English: B. Columbia	0.289	−1.356***	0.265
	(0.444)	(0.287)	(0.509)
Family Wealth: Quebec	−0.064	−0.752***	−0.173
	(0.132)	(0.102)	(0.167)
Family Wealth: Saskatchewan	−0.153	−0.129	0.070
	(0.165)	(0.198)	(0.206)
Family Wealth: B.Columbia	−0.160	−0.709***	0.417**
	(0.166)	(0.147)	(0.200)
Constant	−7.067***	−3.636***	−6.228***
	(0.406)	(0.252)	(0.403)
Akaike Inf. Crit.	10,458.460	10,458.460	10,458.460

Note: *p<0.1; **p<0.05; ***p<0.01

Benchmark: students from Public English schools.

Appendix II

Table 9: *Binary logit: second model estimation*

	Dependent variable:
	Public French
Mother not working	-0.371^{***}
	(0.069)
Two-parent family	0.025
	(0.087)
Siblings	-0.052
	(0.084)
Grandparents	-0.752^{***}
	(0.133)
Cultural Possessions	-0.356^{***}
	(0.034)
Mother Educ. level	0.090^{***}
	(0.028)
Father Educ. level	-0.022
	(0.023)
Family Wealth	0.342^{***}
	(0.113)
Occup. Status	0.011^{***}
	(0.002)
Home Language not English	2.744^{***}
	(0.227)
Alberta	-0.425^{**}
	(0.187)
N. Brunswick	0.333^{**}
	(0.170)
Ontario	1.411^{***}
	(0.134)
Home Language not English: Alberta	-0.778^{***}
	(0.286)
Home Language not English: N. Brunswick	2.225^{***}
	(0.295)
Home Language not English: Ontario	-1.408^{***}
	(0.241)
Family Wealth: Alberta	-0.299^{**}
	(0.142)
Family Wealth: N. Brunswick	-0.407^{***}
	(0.152)
Family Wealth: Ontario	-0.288^{**}
	(0.119)
Constant	-3.388^{***}
	(0.216)
Akaike Inf. Crit.	6,444.261
Note:	*p<0.1; **p<0.05; ***p<0.01

Benchmark: students from Public English schools.